THE ROLE OF CHROMATIN STRUCTURE AND HISTONE MODIFICATIONS IN GENE SILENCING AT THE RIBOSOMAL DNA LOCUS IN *Saccharomyces cerevisiae*

A Dissertation

by

KELLY M. WILLIAMSON

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2011

Major Subject: Biochemistry

THE ROLE OF CHROMATIN STRUCTURE AND HISTONE MODIFICATIONS IN GENE SILENCING AT THE RIBOSOMAL DNA LOCUS IN *Saccharomyces cerevisiae*

A Dissertation

by

KELLY M. WILLIAMSON

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Mary Bryk
Committee Members,	Gary Kunkel
	Carlos Pace
	Dorothy Shippen
Head of Department,	Gregory Reinhart

May 2011

Major Subject: Biochemistry

ABSTRACT

The Role of Chromatin Structure and Histone Modifications in Gene Silencing at the Ribosomal DNA Locus in *Saccharomyces cerevisiae.* (May 2011)

Kelly M. Williamson, B.S., Texas State University

Chair of Advisory Committee: Dr. Mary Bryk

One of the fundamental questions in science is how chromatin transitions from actively transcribed euchromatin to silent heterochromatin, and what factors affect this transition. One area of my research has focused on understanding the differences in the chromatin structure of active and silent regions in the ribosomal DNA locus (rDNA), a heterochromatin region in *S. cerevisiae*. Secondly, I have focused on understanding a histone methyltransferase Set1, which is involved in both euchromatin and heterochromatin regions. To distinguish actively transcribed open regions of chromatin from silent and closed regions of chromatin, we have expressed a DNA methyltransferase *M.CviPI in vivo* to utilize its accessibility to GpC sites. We have used this technique to study changes in nucleosome positioning within the NTS2 region of the rDNA in two cases: as a result of a silencing defect caused by the loss of Sir2, a histone deacetylase involved in silencing at the rDNA, and as an indicator of active transcription by RNA Pol I. Using this technique, we observed differences between open and closed chromatin structure by changes in nucleosome positioning within NTS2. Additionally,

we have observed the presence of bound factors within the 35S rRNA gene promoter that are unique to actively transcribed genes.

The second area of my research focused on the protein methyltransferase Set1 that mono-, di-, and trimethylates lysine 4 of histone H3 (H3K4) utilizing the methyl group from S-adenosyl methionine (SAM). Set1 is part of a multi protein complex called COMPASS (Complex associated with Set1), and is associated with both actively transcribed and silent regions. Thirty mutants of Set1 were made within the SET domain to learn more about the catalytic mechanism of Set1. The crystal structures of human SET domain proteins, as well as sequence alignments and a random mutagenesis of yeast Set1, were used to identify conserved amino acids in the SET domain of Set1. Mutants were analyzed for their effect on histone methylation in vivo, silencing of RNA Pol II transcription within the rDNA, suppression of *ipl1-2*, and COMPASS complex formation. Our results show that trimethylated H3K4 is required for silencing of RNA Pol II transcription at the rDNA. Overall, we have shown the importance of tyrosine residues in SET domain proteins. To summarize, my research has strived to understand chromatin structure and the factors that affect the transition between euchromatin and heterochromatin.

DEDICATION

I dedicate this dissertation to Dr. Mary Bryk and graduate students of the Bryk lab. Hopefully this work is a great stepping off point for many projects and many grants. Both projects discussed in this dissertation have great potential to open avenues of study in many areas, and I hope I have accomplished something that is helpful to future graduate students.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Mary Bryk, and my committee members, Dr. Dorothy Shippen, Dr. Gary Kunkel, and Dr. Nick Pace for their guidance and support throughout the course of this research. I would also like to thank Scott Hoose and Dr. Mike Kladde for their help with *in vivo* expression of *M.CviPI* discussed in Chapter II. Scott Hoose was especially helpful in all dideoxy sequencing performed early on in these experiments that was essential for creating preliminary data. I'd also like to acknowledge Dr. John Mueller for completing the quantitative Westerns measuring histone methylation *in vivo* in Chapter III. Also, I'd like to thank Dr. Larry Dangott for his suggestions and assistance on TAP purification of COMPASS and purification of full length H3 from inclusion bodies.

Thanks also go to previous lab members Chonghua Li and April Stetler, as well as current members Rachel Jordan and Shelley Henderson-Pozzi. I thank all of them for their continual support and advice. Ladies night out is essential to any graduate school survival and Lifesaver café will forever be missed. I'd also like to thank two outstanding previous undergraduate students Morgan Chateau and Victoria Schneider for their help on the Set1 project discussed in Chapter III. Finally, thanks to my family for their encouragement and understanding. Last but not at all least; I'd like to thank my husband for his enormous amount of patience.

NOMENCLATURE

ARS	autonomous replicating sequence
CF	core factor
CLIP	chromosome linkage inner nuclear membrane proteins
COMPASS	complex associated with Set1
FEAR	cdc fourteen early anaphase release
HATs	histone acetyltransferases
HDACs	histone deacetylases
MEN	mitotic exit network
MLL	mixed lineage leukemia
MNase	micrococcal nuclease
Net1	nucleolar silencing establishing factor and telophase regulator
NTS	nontranscribed spacer
Pol	polymerase
rDNA	ribosomal DNA
RENT	regulator of nucleolar silencing and telophase exit
SET	suvar 3-9, enhancer of zeste, trithorax
SIR	silent information regulators
TBP	TATA-binding protein
UAF	upstream activation factor

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

One of the most fundamental purposes of a cell is to properly maintain, store and propagate genetic information. Eukaryotic cells package genetic information in the form of chromosomes made up of DNA, histones and various other proteins that regulate and maintain chromatin structure. Within chromosomes DNA is wrapped around a histone octamer of H2A, H2B, H3 and H4 forming a nucleosome (Finch *et al.*, 1977; Richmond *et al.*, 1984; Luger *et al.*, 1997; Richmond and Davey, 2003; Wood *et al.*, 2005). Neighboring nucleosomes connected by DNA acquire a 'beads on a string' structure, and through interactions between neighboring and distant nucleosomes result in increasing levels of complexity to ultimately shape an entire chromosome (Figure 1) (Felsenfeld and Groudine, 2003).

Chromatin structure can be categorized based on the level of gene expression. Transcriptionally active euchromatin contains an open, diffuse chromatin structure with actively transcribed genes, while transcriptionally silent heterochromatin contains a tightly compacted, closed chromatin structure that functions as a barrier to transcription, replication, and recombination factors (Frenster *et al.*, 1963; Littau *et al.*, 1964; Grewal and Moazed, 2003; Holmquist and Ashley, 2006). Early studies used electron microscopy to directly visualize nuclei purified from calf thymus, identifying two types of chromatin structure: condensed, dark regions of chromatin and open diffuse regions of

---

This dissertation follows the style of Molecular Biology of the Cell.

chromatin that were found to be sites of active transcription (Frenster *et al.*, 1963; Littau *et al.*, 1964). These early studies proposed the idea that actively transcribed euchromatin would contain a low level of histones when compared to inactive heterochromatin and that the majority of transcription would occur in open chromatin regions. Furthermore, they conclude that the majority of chromatin is silenced.

Silencing of transcription is an essential part of any organism's ability to control gene expression. Regions of silent, condensed chromatin form through the association of protein factors that spread throughout a chromosomal domain silencing transcription. The silencing of specific regions of chromatin, rather than the repression of genes in a promoter-dependent fashion, is defined as gene silencing (Moazed, 2001). In our model system *Saccharomyces cerevisiae* gene silencing occurs in three regions: telomeres located at the end of chromosomes, the ribosomal DNA array located on chromosome XII, and the silent mating-type loci located on chromosome III. Though each region utilizes many different proteins to restrict gene expression, members of the Silent Information Regulators, or SIR proteins, can be found at each of these silent regions (Perrod and Gasser, 2003; Rusche *et al.*, 2003). Once established, euchromatin and heterochromatin regions are maintained throughout the following generations. This inheritance of chromatin states is also mirrored by the cell propagating certain patterns of histone modifications, and by inheriting associated chromatin regulators such as chromatin activators or silencing complexes bound to chromatin. In our lab we focus on the ribosomal DNA locus in *S. cerevisiae* to study the role of histones and post-translational modifications of histones in gene silencing.

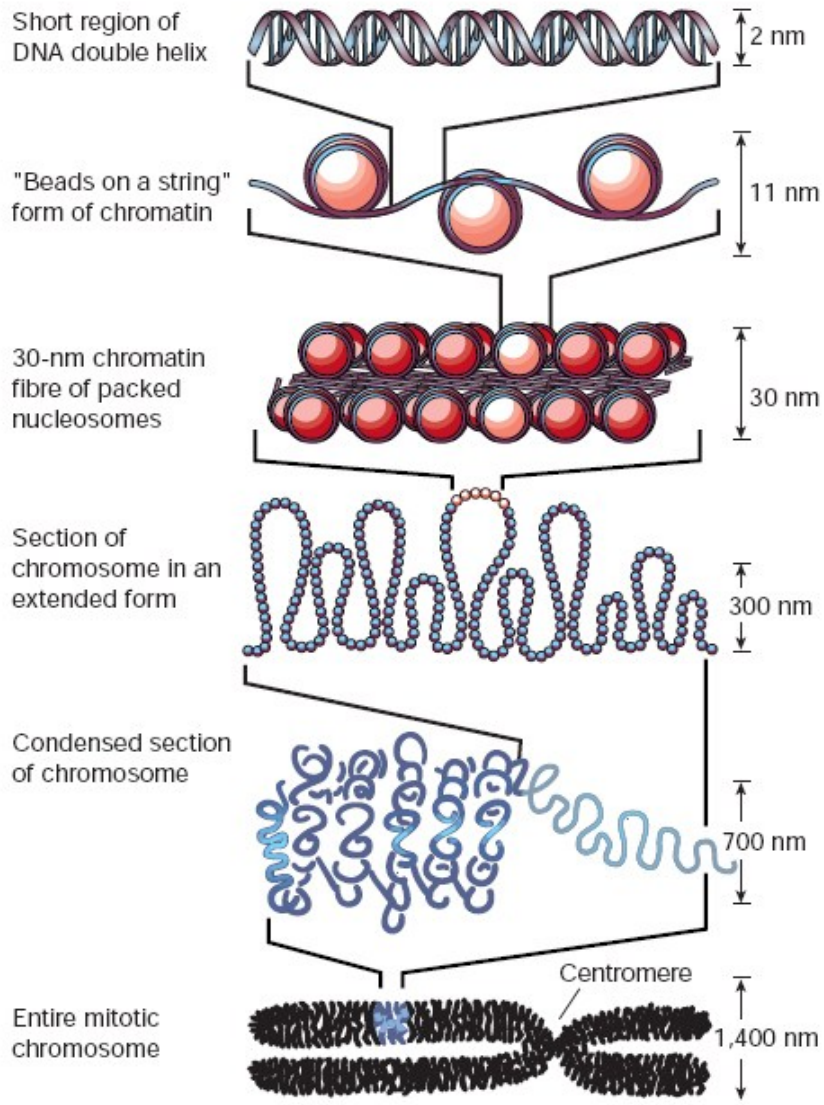


Figure 1. Structure and formation of chromosomes[1]**.** A 2nm DNA strand wraps twice around a histone octamer of H2A, H2B, H3 and H4 forming a nucleosome. This simple structure then folds upon itself until it ultimately forms an entire chromosome.

[1] Adapted by permission from Macmillan Publishers Ltd from "Controlling the double helix" by Gary Felsenfeld and Mark Groudine, (2003), Nature, 421, 448-453, Copyright 2003 by Nature Publishing Group.

All four core histones (H2A, H2B, H3 and H4) have been studied for their role in euchromatin and heterochromatin. Both H3 and H4 play a role in silencing at all three heterochromatin regions in *S. cerevisiae* by maintaining a closed chromatin structure and thus repressing transcription (Kayne *et al.*, 1988; Aparicio *et al.*, 1991; Park *et al.*, 2002; Sperling and Grunstein, 2009). These experiments identified specific amino acid residues on histone H3 and H4 that are required for silencing at heterochromatin regions, and these residues in some cases are well known to undergo post-translational modification (Park *et al.*, 2002; Hyland *et al.*, 2005). Through a combination of truncations or complete deletions of the N-terminus as well as site-specific mutagenesis, a wealth of information exists on the role of histones on chromatin structure and in repression of transcription at heterochromatic regions. The post-translation modification of histones is a strong indicator of euchromatin and heterochromatin regions. Histones can be modified in various ways, most commonly by acetylation, methylation and ubiquitylation. Different combinations of post-translation modifications of histones are thought to form a histone code (Jenuwein and Allis, 2001; Berger, 2002).

There are two ways histone modifications are thought to affect chromatin. Simply, addition of modifications can directly affect chromatin structure by masking the charge of residues present on histone tails. This allows the DNA to have a looser association with histones and thus allow a more open chromatin structure. Modifications can also serve to create a code that can be read by chromatin-associating proteins that will regulate the transition between heterochromatin and euchromatin. By 'reading' a histone code chromatin-associating proteins can affect gene expression. It is believed

that these chromatin-associating proteins bind different combinations of modifications and help recruit other chromatin regulators that will ultimately determine the level of gene expression (Quina *et al.*, 2006). Research in the area of histone modifications has branched out into what is termed 'cross talk', determining how different modifications affect each other. In general, each type of modification is associated with a different role in transcription. Acetylation is associated with actively transcribed regions, while both methylation and ubiquitylation are associated with activation and repression of transcription. In defining the role of histone modifications, it is the location of the mark on a specific amino acid residue(s) in either euchromatin or heterochromatin and the presence or absence of other marks that can influence the final effect of histone modifications on a region of chromatin.

Histone acetylation, the most studied modification, is characterized by histone acetyltransferases (HATs) and histone deacetylases (HDACs) which add and remove acetyl groups to and from the ε-amino group of lysine residues (Yang, 2004; Shahbazian and Grunstein, 2007). Acetylation is associated with activation of transcription, with hyperacetylated regions found in euchromatin, forming an open chromatin structure and increasing gene expression, while hypoacetylated regions are associated with heterochromatin (Braunstein *et al.*, 1993; Hecht *et al.*, 1995; Braunstein *et al.*, 1996; Suka *et al.*, 2001; Bryk *et al.*, 2002; Li *et al.*, 2006a). Initial studies proposed that the acetyl group masks the positive charge of lysine residues, disrupting the interaction of histone tails with the negatively charged DNA backbone thus opening the chromatin to transcription, replication, and recombination factors (Allfrey *et al.*, 1964; Allfrey and

Mirsky, 1964; Vidali *et al.*, 1968; DeLange *et al.*, 1969). Of these early experiments, Allfrey *et al*. was the first to connect acetylated lysines as well as methylated lysines to increased transcription by studying their effect on *in vitro* transcription assays with purified calf thymus nuclei (Allfrey *et al.*, 1964; Allfrey and Mirsky, 1964). Sir2 is a histone deacetylase that plays an important role in heterochromatin formation and maintenance, transcriptional silencing of RNA Pol II, and repression of recombination in *S. cerevisiae.* These and other topics on Sir2 will be discussed in further detail below.

The effect of histone methylation on gene expression depends on the particular lysine residue that is methylated. Histone methylation is an interesting modification because of its dual nature in euchromatin and heterochromatin regions in *Saccharomyces cerevisiae* and higher eukaryotes. Most histone methylation of lysine residues is carried out by proteins that contain a very specific and highly conserved SET domain. The SET domain name itself refers to its origins in *Drosophila melanogaster,* originally recognized in proteins encoded by the *Drosophila* genes Suvar 3-9, Enhancer of zeste, and Trithorax, all of which play some role in regulating transcription and were identified for their similarity to the human histone methyltransferase *MLL1* gene (Jones and Gelbart, 1993; Tschiersch *et al.*, 1994). The role of the SET domain was first defined by the discovery of human Suv39H1 and mouse Suv39h1 as histone methyltransferases that methylate lysine 9 of histone H3 and function exclusively in heterochromatin regions (Rea *et al.*, 2000). SET domain proteins have been found in a wide range of organisms from simple model systems all the way to mouse and human systems. SET domain proteins utilize S-adenosyl methionine (SAM), which provides the methyl group to SET

domain proteins to methylate lysine residues. SAM is formed by AdoMet synthetase which transfers the adenosyl part of ATP to methionine, and is employed in all methylation reactions, for histones and other proteins as well as DNA methylation. Once SAM has been utilized by these proteins, the product S-adenosyl homocysteine is produced (Chiang *et al.*, 1996). Methylation was once considered to be a permanent mark until the discovery of two classes of demethylases, LSD (Shi *et al.*, 2004) and Jumonji (Tu *et al.*, 2007). Histone demethylases are discussed in greater in Appendix B.

Lysine residues can be methylated up to three times by histone methyltransferases (HMTases). In *S. cerevisiae*, lysine 4, 36, and 79 of histone H3 are methylated by Set1, Set2, and Dot1, respectively. Methylation of lysine 4 of histone H3 (H3K4) by Set1 is associated with initiation of transcription in actively transcribed genes, as well as silent chromatin. Methylation of lysine 36 of histone H3 by Set2 plays a role in actively transcribed genes in transcriptional elongation. Methylation of lysine 79 of histone H3 by Dot1 (Disruptor of Telomeric silencing) is associated with silencing of transcription at telomeres (Briggs *et al.*, 2001; Roguev *et al.*, 2001; Strahl *et al.*, 2002; van Leeuwen *et al.*, 2002). Like methylation, ubiquitylation is a mark associated with active and silent genes, as ubiquitylation of lysine 123 of histone H2B is required for di- and trimethylation of H3K4 to occur (Dover *et al.*, 2002; Ng *et al.*, 2003a; Shahbazian *et al.*, 2005). In humans and higher eukaryotes, methylation of H3K4 is solely associated with activation of transcription while methylation of lysine 9 of histone H3, a mark absent from *S. cerevisiae*, is associated with silencing of transcription in higher eukaryotes.

In humans, there are six methyltransferases that act on lysine 4 of histone H3: the MLL family, Set1A/Set1B methyltransferases, Set7/9, Ash1L, Smyd3, and Meisetz Human H3K4 methyltransferases are discussed in Appendix A. In yeast there is only one, making it a much simpler model system for the study of H3K4 methylation. Table 1 summarizes the methyltransferases mentioned in this section. In yeast methylation of H3K4 is carried out by Set1 (Briggs *et al.*, 2001), a histone methyltransferase that is part of a multi-subunit complex called COMPASS (Complex associated with Set1) (Miller *et al.*, 2001; Roguev *et al.*, 2001). COMPASS contains eight subunits including: Set1, Bre2, Swd1, Spp1, Swd2, Swd3, Sdc1, and Shg1. These subunits all contain conserved domains that are found in other histone methyltransferase complexes, like the human MLL family complexes and human Set1A/Set1B complexes. Although yeast Set1 is predominantly studied for its role in actively transcribed genes, it was first discovered for its role in silencing transcription at telomeres and the silent mating type loci (Nislow *et al.*, 1997; Krogan *et al.*, 2002) and then only later found to silence transcription at the rDNA (Briggs *et al.*, 2001; Bryk *et al.*, 2002). At the rDNA, cells with a defect in silencing due to loss of Sir2 exhibit an increase in trimethylation of H3K4 (Li *et al.*, 2006a) suggesting a model where low levels of methylated histones are required for silencing of transcription at the rDNA.

Table 1. Summary of histone methyltransferases.

Model system	Histone Methyltransferase	Target
Saccharomyces cerevisiae	COMPASS	H3K4
	Set2	H3K36
	Dot1	H3K79
Mouse	Suv39h1	H3K9
Human	MLL1-5 complexes	H3K4
	Set1A/Set1B complexes	H3K4
	Set7/9	H3K4
	Ash1L	H3K4
	Smyd3	H3K4
	Meisetz	H3K4
	Suv39H1	H3K9

COMPASS associates with the c-terminal domain of RNA Polymerase II, when phosphorylated at serine 5, in order to trimethylate H3K4 at the promoter regions of highly transcribed genes (Santos-Rosa *et al.*, 2002; Ng *et al.*, 2003b). Phosphorylation of the CTD of RNA Pol II at serine 5 is associated with initiation of transcription at the promoter region of active genes while phosphorylation at serine 2 is associated with elongation. These characteristics of CTD phosphorylation are conserved from yeast to humans (Morris *et al.*, 2005). COMPASS associates with RNA Pol II through the Paf1 (Polymerase associated factors) complex (Krogan *et al.*, 2003), a complex conserved from yeast all the way to humans with a role in ubiquitylation of histone H2B (Jaehning, 2010). Paf1c serves to recruit methyltransferases to the transcriptional machinery and therefore has a role in initiation of transcription (recruitment of yeast Set1 COMPASS) and in elongation of transcription (recruitment of yeast Set2). Paf1c also serves to recruit proteins important in ubiquitylation of H2B (Jaehning, 2010). Di- and

trimethylation of H3K4 by Set1 is dependent upon ubiquitylation of lysine 123 of histone H2B by Rad6 ubiquitin conjugating enzyme and by Bre1 ubiquitin ligase (hereto after referred to as Rad6/Bre1) (Dover *et al.*, 2002; Shahbazian *et al.*, 2005). Histone methylation of H3K4 is distributed in a very distinct pattern along actively transcribed genes (Santos-Rosa *et al.*, 2002), and this pattern is universal in all systems. Trimethylation of H3K4 peaks at the promoter region, dimethylation is in the middle of the gene, while monomethylation is found at the end.

As mentioned above, COMPASS contains eight subunits with conserved domains. Set1 contains the catalytic SET domain and two amino-terminal RNA-recognition motifs (RRM). The RRM domains of Set1 (RRM1 and RRM2) bind single stranded nucleic acids *in vitro* (Krajewski *et al.*, 2005) but this binding does not affect H3K4 methylation (Tresaugues *et al.*, 2006). Deletion of both RRM1 and RRM2 causes a loss of trimethylated H3K4 only and does not affect dimethylated H3K4 (Schlichter and Cairns, 2005). Recent work has shown that Set1 associates with its own mRNA along with Swd1, Spp1, and Shg1 to form SET1RC (SET1 mRNA-associated complex), but this association does not rely on either of the RRM domains of Set1 (Halbach *et al.*, 2009). The purpose of the RRM domains of Set1 remains unknown. SET1RC is believed to form co-translationally; the complex associates with the Set1 amino acid strand as it is translated from the *SET1* mRNA and thus couples synthesis of Set1 with formation of the complex.

Spp1 is a subunit of COMPASS that contains a PHD domain that specifically recognizes di- and trimethylated H3K4 (Shi *et al.*, 2007), and contains a DNA binding

domain that tethers COMPASS to chromatin (Murton *et al.*, 2010). Swd1, Swd2, and Swd3 all contain WD40 domains, and are homologues to the human WD40-containing WDR5 that binds dimethylated H3K4 and presents this modified lysine for trimethylation to H3K4 methyltransferases in humans (Wysocka *et al.*, 2005b). Swd1 weakly binds dimethylated H3K4 (Murton *et al.*, 2010). Swd2 is ubiquitylated at lysine 68 and 69 by Rad6/Bre1 and this modification controls whether or not Spp1 is recruited by COMPASS and thus whether H3K4 is trimethylated; however, it does not control the overall stability of COMPASS as mono- and dimethylation of H3K4 still occurs (Vitaliano-Prunier *et al.*, 2008). These domains are mirrored in human methyltransferases and are discussed in Appendix A.

Interestingly enough, SET domain proteins have also been found to methylate non-histone proteins at lysine residues forming new roles of SET domain proteins (Huang and Berger, 2008). Alternative substrates of human SET domain methyltransferases Set7/9 are discussed in Appendix C. Yeast Set1 dimethylates Dam1, a protein involved in chromosome segregation and proper microtubule attachment. Methylation of Dam1 is part of the anaphase checkpoint during the cell cycle. Dam1 is part of a large ten subunit complex called the Dam1 complex (Hofmann *et al.*, 1998; Cheeseman *et al.*, 2001; Enquist-Newman *et al.*, 2001). Approximately 16-20 Dam1 complexes form a single Dam1 ring that encircles microtubules, sliding along the microtubule surface creating energy that is used to power chromosome movement and thus segregation of chromosomes during anaphase (Westermann *et al.*, 2005; Westermann *et al.*, 2006; Wang *et al.*, 2007a; Westermann *et al.*, 2007). The interaction

between the kinetochore complexes located at the centromere of chromosomes and microtubules is controlled with phosphorylation by Ipl1 and dephosphorylation by Glc7 (Pinsky *et al.*, 2006a; Pinsky *et al.*, 2006b; Westermann *et al.*, 2007). Ipl1 is an aurora kinase that ensures proper kinetochore attachment and chromosome segregation by resolving incorrect attachments of microtubules to the kinetochore. Ipl1 phosphorylation of multiple targets causes improperly attached microtubules to detach from the kinetochore, allowing for a second attempt to properly attach to the kinetochore. This is part of the spindle checkpoint will continue until proper attachment occurs and chromosome segregation can proceed (Pinsky *et al.*, 2006b).

Ipl1 phosphorylates Dam1 in response to improperly attached microtubules (Cheeseman *et al.*, 2002), while Set1 methylates Dam1 in response to properly attached microtubules, blocking phosphorylation by Ipl1 (Zhang *et al.*, 2005). In other words, Set1 methylation antagonizes Ipl1 phosphorylation. Dam1 is dimethylated by Set1 at an internal lysine residue (K233) flanked by serine residues that are sites of phosphorylation by Ipl1. Methylation of Dam1 at K233 by Set1 physically blocks the phosphorylation by Ipl1 of those neighboring serine residues (Zhang *et al.*, 2005). This new role for Set1 methylation was discovered using an *ipl1-2* temperature sensitive (ts) yeast strain. At a restrictive temperature of 30°C *ipl1-2* is inactive and cell growth is inhibited, while in yeast strains combining *ipl1-2* and *set1Δ*, normal cell growth is restored meaning deletion of *SET1* suppresses the growth defect of *ipl1-2* and therefore Set1 has some role in Ipl1 function (Zhang *et al.*, 2005). Methylation of Dam1 by Set1

inhibits continued phosphorylation by Ipl1, allowing segregation of chromosomes during anaphase.

HETEROCHROMATIN REGIONS IN *Saccharomyces cerevisiae*

Silencing of transcription is an essential part of any organism's ability to control gene expression. In our model system *Saccharomyces cerevisiae*, commonly referred to as baker's yeast, transcriptional silencing occurs in three regions: telomeres located at the end of chromosomes, the ribosomal DNA array located on chromosome XII, and the silent mating-type loci located on chromosome III. Chromosome III consists of two silenced loci *HML* and *HMR* on either side of the centromere, near the telomeres of chromosome III with a mating type locus (*MAT*) in between. Yeast can function as haploids or diploids, with haploids occurring in one of two mating types: α and $\underline{a}$. The mating type is determined by the genetic information expressed at the *MAT* locus, either *MAT*α *or MAT*$\underline{a}$ (Klar *et al.*, 1981; Nasmyth *et al.*, 1981). *HML* and *HMR* carry extra silenced copies of α and $\underline{a}$, with α at *HML* and $\underline{a}$ at *HMR*. Chromatin within the silent mating type loci *HML* and *HMR* represses transcription, ensuring that only the information present at the *MAT* locus is expressed (Rusche *et al.*, 2003). The genetic information expressed at the *MAT* locus can change due to *HO* endonuclease, which cuts within the *MAT* locus causing a double-stranded break that is repaired through homologous recombination between one of the silenced *HML* or *HMR* loci and identical sequences within the *MAT* locus (Rusche *et al.*, 2003). The gene encoding *HO*

endonuclease is deleted in lab strains to maintain cells as one mating type and strains are unable to switch between the two.

Though each region utilizes many different proteins to restrict gene expression, members of the Silent Information Regulators, or SIR proteins can be found at each of these silent regions (Perrod and Gasser, 2003; Rusche *et al.*, 2003). The histone deacetylase (HDAC) Sir2 is required for transcriptional silencing of gene expression at all three regions and specifically targets lysine 9, 14, and 56 of histone H3 and lysine 16 of histone H4 (Imai *et al.*, 2000a; Imai *et al.*, 2000b; Xu *et al.*, 2007). In *S. cerevisiae*, silent chromatin is characterized by low levels of acetylated and methylated histones while actively transcribed chromatin is characterized by high levels of acetylated and methylated histones (Braunstein *et al.*, 1993; Hecht *et al.*, 1995; Braunstein *et al.*, 1996; Suka *et al.*, 2001; Bryk *et al.*, 2002; Li *et al.*, 2006a).

The recruitment and spreading of SIR proteins along chromatin differs between each of the three regions of heterochromatin in *S. cerevisiae.* At both telomeres and the silent mating type loci Sir2, Sir3, and Sir4 interact and spread along both regions to form heterochromatin (Aparicio *et al.*, 1991; Gotta *et al.*, 1997; Hoppe *et al.*, 2002); however, at the rDNA only Sir2 is required for silencing of transcription (Smith and Boeke, 1997; Straight *et al.*, 1999; Hoppe *et al.*, 2002). A cycle of recruitment and deacetylation of histones facilitates spreading along chromatin in just a few steps. At the telomeres, formation of silent chromatin is initiated by the binding of a telomere binding protein called Rap1 (Repressor/Activator Protein 1) to specific sites (Klein *et al.*, 1992; Kyrion *et al.*, 1992; Moretti *et al.*, 1994; Luo *et al.*, 2002). Rap1 and Ku70/80 work together to

specifically recruit Sir4 to telomeres (Mishra and Shore, 1999) followed by the HDAC Sir2 which deacetylates H3 and H4 (Imai *et al.*, 2000a), recruiting Sir3 followed by Sir4, etc. Hypoacetylated H3 and H4 N-terminal tails are the preferred binding site for Sir4 and Sir3, and thus a Sir2-Sir3-Sir4 complex spreads along telomeres forming a silenced, compact chromatin structure (Braunstein *et al.*, 1993; Hecht *et al.*, 1995; Suka *et al.*, 2001; Sperling and Grunstein, 2009). As mentioned above, Dot1 methylates lysine 79 of histone H3 which is important for silencing at telomeres and is believed to specifically target and restrict SIR proteins to heterochromatin regions. With a loss of methylation at lysine 79 of histone H3, SIR proteins spread further along chromatin depleting SIR proteins from heterochromatin and leading to a loss of silencing (van Leeuwen *et al.*, 2002; Verzijlbergen *et al.*, 2009).

The heterochromatin structure of telomeres is stabilized by folding back on itself (de Bruin *et al.*, 2000) upon associating with specific foci along the nuclear periphery containing high levels of silencing proteins (Maillet *et al.*, 1996; Andrulis *et al.*, 1998). Association of telomeres with the nuclear periphery is believed to be important for interaction with specific foci containing the silencing proteins Rap1, Sir2, Sir3, Sir4 and both Ku proteins (Klein *et al.*, 1992; Palladino *et al.*, 1993; Gotta *et al.*, 1996; Maillet *et al.*, 1996; Laroche *et al.*, 1998; Mondoux *et al.*, 2007). Ku70 and Ku80 are believed to function as important anchors of telomeres to these foci at the nuclear periphery (Boulton and Jackson, 1998; Laroche *et al.*, 1998; Andrulis *et al.*, 2002). Deletion of any of the five proteins listed above can result in loss of silencing at telomeres, chromosome instability, and loss of formation of the silencing foci. The current model

suggests that localization of telomeres at the periphery is not required for silencing but is required for interaction with silencing proteins leading to the initiation and thus maintenance of silencing (Mondoux *et al.*, 2007). In other words, once silencing is initiated telomeres can travel away from the periphery and be silent but association with the nuclear periphery is required for interaction with a pool of SIR proteins and others to initiate silencing of gene expression.

The recruitment and deacetylation cycle described above is similar to the spreading of Sir2-Sir3-Sir4 in the silent mating type loci (Moazed, 2001; Rusche *et al.*, 2003). Both of the *HM* loci are flanked by silencer regions that recruit Rap1, the origin recognition complex (ORC) and Abf1. Unlike the other heterochromatin regions, *HM* loci rely on Sir1 to initiate the recruitment of other SIR proteins followed by spreading as described above (Rine and Herskowitz, 1987; Aparicio *et al.*, 1991; Bose *et al.*, 2004). Localization of the *HM* loci to the nuclear periphery has also been found to be important in silencing (Maillet *et al.*, 1996; Andrulis *et al.*, 1998; Andrulis *et al.*, 2002), and as with telomeres it is believed that association with the periphery is important in initiating silencing but is not required for maintaining silencing once established (Gartenberg *et al.*, 2004).

Ribosomal DNA locus

The ribosomal DNA locus consists of a 1-2 Mb rDNA tandem array containing ~150 to 200 copies of a 9.1 kb ribosomal DNA repeat on chromosome XII forming 10% of the genome in *S. cerevisiae* (Petes *et al.*, 1978b; Skryabin *et al.*, 1978; Petes, 1979; Warner, 1999). A single rDNA repeat consists of two transcribed regions, the 35S

rRNA gene and the 5S rRNA gene. The 35S rRNA gene is transcribed by RNA Polymerase I (RNA Pol I) and is the precursor of 25S, 5.8S and 18S rRNAs. The 5S rRNA gene is transcribed by RNA Polymerase III (RNA Pol III). RNA Pol I accounts for up to 60% of the total transcription in *S. cerevisiae* (Warner, 1999). Two nontranscribed spacer regions NTS1 and NTS2 surround the 5S gene (Figure 2a) (Skryabin *et al.*, 1984). The rDNA array compacts along the interior of the nucleus forming a 'cap' structure called the nucleolus. While telomeres serve as a classic example of heterochromatin, the ribosomal DNA locus is best described as facultative heterochromatin; a region that is contains high levels of transcription as well as regions of silenced transcription.

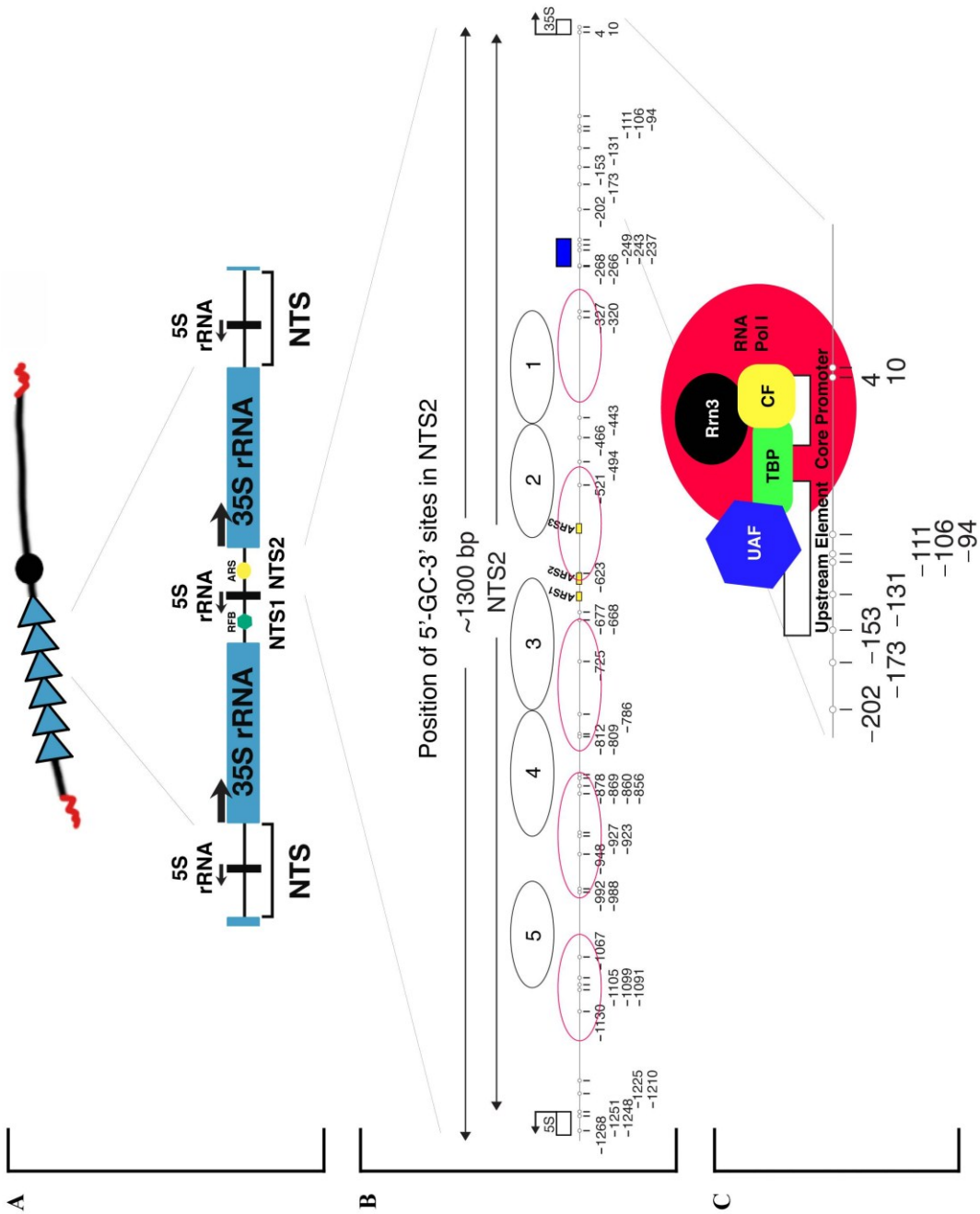
Of the SIR proteins mentioned above, only Sir2 is required for silencing at the rDNA (Gotta *et al.*, 1997; Smith and Boeke, 1997; Straight *et al.*, 1999; Hoppe *et al.*, 2002). Gene silencing was demonstrated at the rDNA by silencing of Ty1 elements as well as other RNA Pol II transcribed genes (Bryk *et al.*, 1997; Smith and Boeke, 1997). Ty1 elements are retrotransposons transcribed by RNA Polymerase II (RNA Pol II) that inserts themselves upstream of tRNA genes, and insertion is dependent upon transcription by RNA Pol III at that tRNA gene (Devine and Boeke, 1996). Silencing at the rDNA is not absolute, instead silencing results in a large decrease but not a complete loss in RNA Pol II transcription. Silencing at the rDNA is a result of a specialized chromatin structure that depends upon Sir2 to prevent transcription by RNA Pol II. Sir2 has a larger effect on the NTS regions than on the transcribed regions of the rDNA (Bryk *et al.*, 1997; Smith and Boeke, 1997). In cells lacking Sir2, there is an increase in

trimethylated H3K4 within NTS2 (Li *et al.*, 2006a). Sir2 also silences small transcripts that naturally occur within the NTS by RNA Pol II transcription (Li *et al.*, 2006a). These transcripts are associated with disrupting the association of cohesin at the rDNA (Kobayashi and Ganley, 2005). The role of cohesin is discussed in further detail below.

Role of NTS1 and NTS2 in DNA replication

Both NTS1 and NTS2 regions play important roles in DNA replication at the rDNA. Autonomous replicating sequence (ARS) elements within NTS2 consist of three 10-11bp sequences that are conserved among rDNA repeats (Miller and Kowalski, 1993) (Figure 2b). DNA replication at the rDNA initiates bidirectionally within NTS2 at conserved ARS elements that function as origins of replication (Skryabin *et al.*, 1984; Brewer and Fangman, 1987; Linskens and Huberman, 1988). Electron microscopy was used to visualize replication bubbles in the NTS opening in a concerted manner in early S-phase downstream of actively transcribed 35S repeats (Saffer and Miller, 1986). Only the replication forks traveling towards NTS1 are halted by a replication fork barrier (RFB) preventing RNA Pol I transcriptional machinery on actively transcribed 35S rRNA genes and DNA replication machinery from colliding (Linskens and Huberman, 1988; Kobayashi and Horiuchi, 1996; Kobayashi, 2003). Replication forks travelling in the same direction as RNA Pol I transcription are allowed to pass through the RFB site. The arrest of replication is independent of transcription by RNA Pol I and RNA Pol II and depends solely on proteins associating with the RFB (Brewer *et al.*, 1992). Fob1 (fork blocking less) binds at the RFB at specific sequences within NTS1 and functions to

Figure 2. The Ribosomal DNA locus in *S. cerevisiae.* **(A)** The rDNA is located on chromosome XII with ~150 to 200 copies of a single rDNA repeat. The 35S and 5S rRNA genes are labeled along with the RFB (replication fork barrier) as a green circle in NTS1 and the ARS (autonomous replicating sequence) as a yellow circle in NTS2. **(B)** NTS2 is expanded showing the positioning of five nucleosomes in black as published by Vogelauer *et al.* 1998 and Li *et al.* 2006. Red circles represent positioning of the five nucleosomes based on data shown in Chapter II of this manuscript. All 5'-GC-3' sites are shown by slash marks in relation to the initiation site of RNA Polymerase I (+1) and are discussed in greater detail in Chapter II. Conserved ARS1, 2, and 3 as published by Miller and Kowalski, 1993 are represented by three small yellow boxes between nucleosome 2 and 3. A blue box at ~ -250 represents a region of high methylation and is discussed in greater detail in Chapter II. **(C)** The promoter region of RNA Polymerase I is further expanded showing the core RNA Pol I complex of UAF (upstream binding factor) in blue, TBP (TATA binding protein) in green, and CF (Core factor) in yellow.


A
5S rRNA
NTS
35S rRNA
RFB
ARS
NTS1
NTS2
B
Position of 5'-GC-3' sites in NTS2
~1300 bp
NTS2
5S
35S
ARS1
ARS2
ARS3
C
Rrn3
CF
TBP
UAF
RNA Pol I
Upstream Element
Core Promoter

halt DNA replication and RNA Pol I transcriptional machinery (Kobayashi *et al.*, 1992; Kobayashi and Horiuchi, 1996). Once the fork is arrested within NTS1, DNA replication machinery continues traveling in the same direction as RNA Pol I (Brewer and Fangman, 1988; Linskens and Huberman, 1988).

Although each of the ~150 to 200 copies of the rDNA contains an ARS sequence only ~ 20% of those sequences are used to initiate DNA replication (Linskens and Huberman, 1988; Muller *et al.*, 2000; Pasero *et al.*, 2002). Active origins are only initiated downstream of actively transcribed RNA Pol I genes (Muller *et al.*, 2000). In general, ARS elements that are closely spaced (~7.5 kb apart) interfere with each other (Brewer and Fangman, 1994). Single molecule analysis of the rDNA repeat shows active origins cluster within the rDNA and are separated by large gaps of inactive origins greater than 60 kb, and that the majority of all origins that do initiate replication do so in early S phase (Pasero *et al.*, 2002). Origins were directly visualized by stretching strands of the rDNA across slides in a technique called molecular combing in combination with BrdU incorporation at actively replicated regions and specific fluorescent probes. In addition to the role of Sir2 in transcriptional silencing, it also performs a role in repression of origin initiation. Sir2 is known to associate within NTS2 at the ARS element (Gotta *et al.*, 1997; Hoppe *et al.*, 2002). In cells lacking Sir2, origin initiation increases between 60-80% and these newly activated origins originate within the normally inactive 60kb gaps mentioned above (Pasero *et al.*, 2002).

RNA Polymerase I transcriptional machinery

NTS2 is also home to the promoter region of the 35S rRNA gene transcribed by RNA Pol I. The 5S rRNA gene contains an internal promoter within the gene that contains internal control regions (ICRs) which bind the RNA Pol III transcriptional machinery (Lee *et al.*, 1995; Paule and White, 2000). These regions are shown in Figure 2 and are discussed in Chapter II. *In vitro* transcription assays with the rDNA was first performed with extracts from mouse cells, showing that RNA Pol I initiates at the 5' end of the large rRNA gene (Grummt, 1981). RNA Pol I activity was more robust in extracts from actively dividing cells than from cells that have reached senescence. Up to this point nothing was known about the proteins required for RNA Pol I initiation and transcription. To identify factors important for RNA Pol I transcription in *S. cerevisiae*, synthesis of rRNA genes must be maintained. Thus, a system was created in which an RNA Pol II promoter *GAL7* was fused at the 5' end of the 35S rRNA gene. In this system, 35S rRNA transcription is maintained by RNA Pol II allowing researchers to screen for mutants that affect RNA Pol I and still preserve cell viability (Nogi *et al.*, 1991a; Nogi *et al.*, 1991b). From this initial study three transcription factors were discovered: UAF (upstream activation factor), CF (core factor), and Rrn3p (Figure 2c).

The RNA Pol I promoter region within NTS2 contains two domains, an upstream element that stretches from -155 to -60 and a core promoter that stretches from -38 to +5 (Musters *et al.*, 1989; Kulkens *et al.*, 1991; Choe *et al.*, 1992) (Figure 2c). Using an *in vitro* system, transcription is detected from the core promoter but is increased ten-fold when the upstream element is included. Thus, the upstream element is required for a

higher level of transcription and the core promoter is required at minimum for RNA Pol I transcription (Kulkens *et al.*, 1991; Choe *et al.*, 1992). UAF, CF, Rrn3, TBP and RNA Pol I are known to bind within the upstream element and core factor regions forming the RNA Pol I promoter region (Figure 2c). The binding and formation of the RNA Pol I transcriptional machinery leads to an actively transcribed 35S rRNA gene.

The Upstream Activation Factor (UAF) is comprised of Rrn5, Rrn9, Rrn10 (Keys *et al.*, 1996), Uaf30 (Siddiqi *et al.*, 2001b), and histones H3 and H4 (Keener *et al.*, 1997) and binds specifically to the upstream element in NTS2 (Keys *et al.*, 1996) (Figure 2c). Without the upstream element or without UAF, transcription by RNA Pol I does occur however only at basal levels. Cells lacking Rrn5, Rrn9, Rrn10, and Uaf30 are extremely sick but are viable, and the level of RNA Pol I transcription is reduced. This suggests that UAF is not essential for RNA Pol I transcription but is required for the high levels of RNA Pol I transcription needed for normal growth. In cells lacking the Uaf30 subunit RNA Pol I transcription at the rDNA is decreased by ~70% (Hontz *et al.*, 2008). The amount of RNA Pol I transcriptional machinery more than doubles on actively transcribed rDNA to compensate for this decrease. The Uaf30 subunit is also required for the UAF complex to bind to the Upstream Element within the RNA Pol I promoter (Hontz *et al.*, 2008). UAF initiates the formation of a pre-initiation complex bound at the upstream element (Keys *et al.*, 1996). UAF is also comprised of histones H3 and H4; accounting for the ability of UAF to bind strongly to the upstream element region in such a way that it rarely binds a second template (Keener *et al.*, 1997). In cells lacking H3 or H4 ~90% of rRNA synthesis is inhibited resulting from a three-fold decrease in

the association of UAF with RNA Pol I promoter region (Tongaonkar *et al.*, 2005). UAF is tightly bound at the RNA Pol I promoter region through its interactions with H3 and H4 whether or not active transcription by RNA Pol I takes place (Keener *et al.*, 1997; Claypool *et al.*, 2004). UAF remains bound to the 35S promoter through multiple rounds of transcription by RNA Pol I (Keys *et al.*, 1996; Aprikian *et al.*, 2001; Claypool *et al.*, 2004).

Core factor (CF) is composed of Rrn6, Rrn7, and Rrn11 proteins (Keys *et al.*, 1994; Lalo *et al.*, 1996; Lin *et al.*, 1996). All three subunits are essential for RNA Pol I transcription and cell viability, and interact with the TATA-binding protein (TBP) (Lalo *et al.*, 1996; Lin *et al.*, 1996). The name core factor comes from the requirement of the complex of Rrn6, 7 and 11 in directing RNA Pol I transcription from the core promoter region (Lin *et al.*, 1996) (Figure 2c). Without Rrn11 RNA Pol I is unable to initiate transcription from neither the full promoter region nor from the core promoter only. A complete complex of CF in addition to, at minimum, the core promoter region is required for RNA Pol I transcription (Lin *et al.*, 1996).

Rrn3 is a protein that physically interacts with RNA Pol I to initiate transcription and then dissociates during elongation (Yamamoto *et al.*, 1996; Milkereit and Tschochner, 1998). The bulk of Rrn3 exists as a free protein while a smaller portion forms an association with RNA Pol I to initiate transcription (Milkereit *et al.*, 1997). Rrn3 interacts not only with RNA Pol I through its A43 subunit but also with Rrn6, a subunit of CF. In an *in vitro* assay RNA Pol I is able bind to the promoter without Rrn3, however the complex is transcriptionally inactive (Aprikian *et al.*, 2001). Rrn3 is

believed to bridge the interaction between RNA Pol I and CF to stimulate binding of RNA Pol I and CF to the promoter to initiate transcription (Peyroche *et al.*, 2000) (Figure 2c).

The TATA-binding protein (TBP) is a transcription factor that is required for initiation of transcription by all three RNA Polymerases in eukaryotic cells. The role of TBP in RNA Pol II transcription is well documented. Briefly, TBP is known for binding directly to a conserved TATA box sequence within the promoter regions to further recruit additional transcription factors. Conversely, the RNA Pol I promoter region lacks a TATA box sequence and is instead recruited through protein interactions. TBP specifically interacts with Rrn6 of CF and Rrn9 of UAF to stimulate transcription, and UAF utilizes its interaction with TBP to recruit CF (Steffan *et al.*, 1996) (Figure 2c). Basal levels of transcription only require RNA Pol I, Rrn3, and CF. It is the association of TBP and UAF at the Upstream Element that elevates the level of RNA Pol I transcription (Keener *et al.*, 1998).

In summary, RNA Pol I utilizes UAF, CF, Rrn3, and TBP to form an initiating complex. UAF binds to the Upstream Element to begin assembly of the RNA Pol I initiation complex (Keys *et al.*, 1996; Steffan *et al.*, 1996). UAF can bind the Upstream Element in the absence of CF and RNA Pol I (Aprikian *et al.*, 2001). Next, Core Factor and RNA Pol I bound with Rrn3 are recruited to the promoter region. This recruitment requires TBP and is dependent upon UAF bound to the Upstream Element, however it is unknown whether TBP is recruited by UAF or enters with CF and RNA Pol I with Rrn3 (Aprikian *et al.*, 2001; Siddiqi *et al.*, 2001a). The promoter region contains poised

transcriptional machinery that allows the cell to rapidly respond to a demand for increased ribosome synthesis. Once the pre-initiation complex is formed transcription begins and enters elongation. RNA Pol I begins transcribing the 35S rRNA gene while TBP, CF, and UAF remain bound to the promoter region (Aprikian *et al.*, 2001; Bier *et al.*, 2004; Claypool *et al.*, 2004).

The relationship between RNA Pol I and Pol II transcription

The rDNA contains a specialized chromatin structure that allows transcription by RNA Pol I and III but silences RNA Pol II in regions that undergo high levels of RNA Pol I and Pol III transcription. Silencing spreads along the rDNA array in one direction towards the centromere (Figure 2a). This spreading of silenced chromatin towards the centromere is dependent upon Sir2 and RNA Pol I transcription (Buck *et al.*, 2002). Deletion of the 35S rRNA core promoter abolishes silencing of RNA Pol II transcription (Buck *et al.*, 2002; Cioci *et al.*, 2003), and if the orientation of the 35S rRNA gene is flipped silencing spreads to the right, toward the centromere (Buck *et al.*, 2002). Silencing of an RNA Pol II gene specifically takes place in actively transcribed 35S rRNA repeats only and silencing of RNA Pol II does not occur in inactive 35S rRNA repeats (Cioci *et al.*, 2003). In a yeast strain with 25 copies of the rDNA repeat, where all 35S rRNA genes are actively transcribed by RNA Pol I, silencing of RNA Pol II transcription is stronger than in wild-type cells with 150 to 200 copies of the rDNA repeat. To summarize, silencing of RNA Pol II transcription requires high levels of RNA Pol I transcription found in active repeats (Buck *et al.*, 2002; Cioci *et al.*, 2003). This is referred to as the reciprocal silencing model.

Chromatin structure and nucleosome positioning

There are three techniques commonly used to study chromatin structure and nucleosome positioning at the rDNA: *in vitro* footprinting techniques such as micrococcal nuclease (MNase) accessibility, *in vitro* psoralen crosslinking, and *in vivo* DNA methyltransferase accessibility (Figure 3). These techniques can each provide data on where factors such as nucleosomes or other DNA binding factors associate with chromatin, and whether chromatin is in an open or closed conformation. Micrococcal nuclease (MNase) is a non-specific endo-exonuclease that is ~17kDa. MNase assays use purified nuclei or permeabilized cells *in vitro* with low levels of enzyme to create partial digestion of the chromatin. These chromatin fragments are then separated by size and probed for an area of interest to reveal a footprint of where factors bind along a specific region (Figure 3a). Psoralen is a three-ring compound produced in plants that when incubated with chromatin will intercalate with increased frequency within open areas of chromatin rather than closed areas of chromatin. Once treated with UV light, psoralen covalently bonds with the DNA in these regions. DNA is purified, cut with restriction endonucleases, and fragments are separated by size. Chromatin that was in an open conformation by active transcription will contain more psoralen and thus migrate slower than chromatin in a closed conformation containing a low amount of psoralen. This technique was used successfully in studying open and closed chromatin at the rDNA, showing that 50% of rDNA repeats contain open and actively transcribed chromatin while the other 50% contain closed and silent chromatin (Dammann *et al.*, 1993)(Figure 3b).

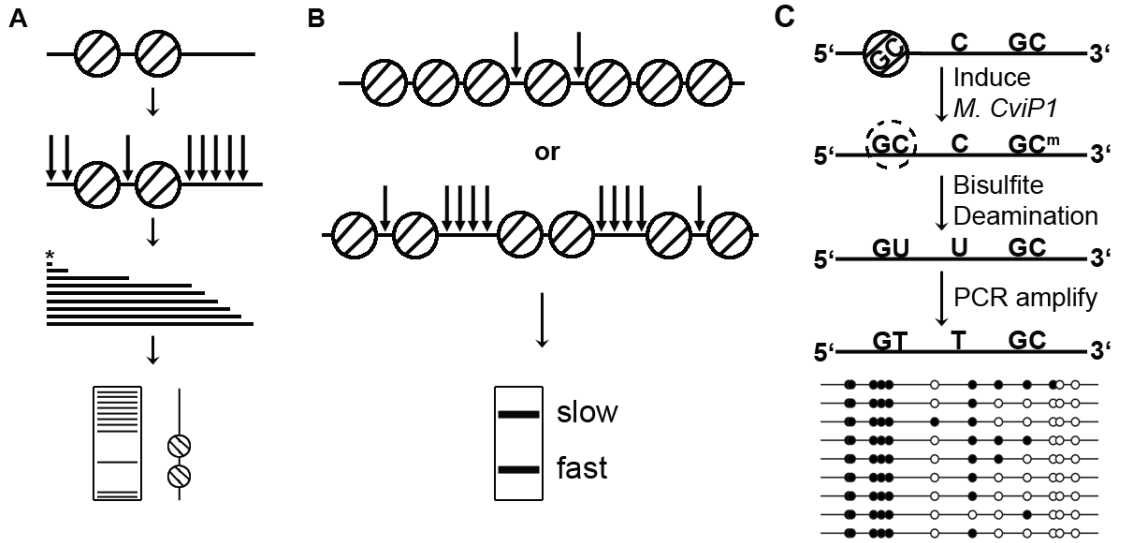


Figure 3. Overview of commonly used techniques in chromatin structure and nucleosome positioning. (A) A Micrococcal nuclease assay. These experiments have typically been used to study nucleosome positioning. In this experiment MNase will cut DNA at areas that are accessible. Areas protected by nucleosomes or other DNA binding proteins are not as accessible and are therefore protected. A partial digestion of the DNA will create fragments of different lengths. By probing for an area of interest, nucleosomes and other DNA binding factors are seen as a footprint. **(B)** Psoralen crosslinking assay. These experiments have been used in the past to study active and inactive repeats of the rDNA (Dammann *et al.*, 1993). In these experiments, actively transcribed repeats that have an open chromatin structure will associate with more psoralen. Thus when probing specifically for the rDNA, two fragments are formed. A slow moving fragment contains more psoralen and thus represents open and actively transcribed repeats while a fast moving fragment contains less psoralen and thus represents closed and silent repeats. **(C)** *In vivo* induction of *M.CviPI*. In these experiments as referenced in the text, the methyltransferase is expressed *in vivo* and recognizes accessible 5'-GpC-3' sites in open chromatin structure. Sites that are accessible are methylated and protected from conversion to uracil. This methylated 5'-GpC^{M}-3' sequence becomes a mark of open, accessible chromatin and can be analyzed by sequence analysis in an area of interest. A common way of displaying data is with a 'dot diagram' as explained in the text.

DNA methyltransferase accessibility is based on experiments that show positioned nucleosomes can inhibit methylation of DNA (Kladde and Simpson, 1994) and this technique can be adapted to map nucleosome positioning and thus chromatin structure *in vivo* (Kladde and Simpson, 1996). Various DNA methyltransferases have been adapted for this procedure, either those expressed endogenously in human cells (Li *et al.*, 2006b) or those artificially inserted into the genome in *Saccharomyces cerevisiae* (Kladde *et al.*, 1996). In human systems, DNA methylation occurs naturally, while in yeast a DNA methyltransferase can be artificially inserted into the genome to study nucleosome positioning and chromatin structure. *M.CviPI* is a methyltransferase that has been utilized in our lab and others to study chromatin structure. *M.CviPI* recognizes 5'-GpC-3' sites (GpC sites), methylating the cytosine residue in open chromatin structure. This specificity by *M.CviPI* provides a recognition site approximately every 27 bp, providing good coverage of a region of interest (Carvin *et al.*, 2003). DNA methylation can be analyzed using a chemical treatment called bisulfite deamination, during which cytosine residues that are unmethylated are converted to uracil while those that are methylated (and thus were accessible in an area of open chromatin) remain unchanged. We and others (Pardo *et al.*, 2009) have developed a single molecule technique by expressing *M.CviPI in vivo* and cloning areas of interest into plasmids. Using this technique we can study changes in nucleosome positioning as an indicator of active transcription by RNA Pol I and as a result of the loss of Sir2. Using this technique we can differentiate between open and closed chromatin structure by the presence of bound factors within the 35S rRNA gene promoter region and changes in

nucleosome positions within NTS2. A common way of displaying data is through a 'dot diagram' where methylation sites are drawn to scale. A black dot represents a methylated site, and therefore an area of open chromatin, and a white dot represents an unmethylated site and therefore an area of closed chromatin (Figure 3c).

Transcriptionally active and inactive regions of chromatin can be described by the sort of chromatin structure present at those regions. Active and inactive regions of the rDNA have been distinguished using psoralen to show that at any one time approximately half of all rDNA genes are actively transcribed and the other half are silent. These two populations can be distinguished by focusing on the entire non-transcribed spacer region (Dammann *et al.*, 1993) and depend upon elongating RNA Pol I to establish an open chromatin structure (Dammann *et al.*, 1995). Within the rDNA actively transcribed genes are randomly distributed along the array, although in the nucleus these active genes may cluster together due to folding of the chromatin forming points of active transcription within the nucleolus (Dammann *et al.*, 1995). Open and active repeats are not maintained from one cell cycle to the next, meaning that there seems to be no preference for one repeat or another (Dammann *et al.*, 1995; Lucchini and Sogo, 1995).

Footprinting techniques were used to define the chromatin structure of the NTS regions including the 35S rRNA gene promoter region. In general it is believed that silent areas of chromatin contain a strict nucleosome positioning, compacted to keep DNA binding factors from disrupting positioning and opening chromatin, while actively transcribed regions may contain few to no nucleosomes that lack a strict position.

Actively transcribed rDNA genes contain approximately 50 complexes of RNA Pol I (French *et al.*, 2003) and have an open chromatin structure are characterized by a lack of nucleosomes while inactive rDNA regions contain strictly positioned nucleosomes (Dammann *et al.*, 1993, 1995). Within NTS2 there are five well-positioned nucleosomes while in NTS1 nucleosome positions are not well defined and cannot be detected (Vogelauer *et al.*, 1998; Lucchini *et al.*, 2001) (Figure 2b). The passage of DNA replication machinery also affects nucleosome positioning. Positioned nucleosomes are temporarily displaced but once the machinery has passed, positioned nucleosomes rapidly reform within NTS2 (Lucchini *et al.*, 2001). When replication machinery enters an actively transcribed gene, positioned nucleosomes reform on the resulting daughter strands. Our lab has utilized *M.CviPI* to study nucleosome positioning within NTS2 as an indicator of actively transcribed genes.

The histone deacetylase Sir2 regulates chromatin structure and nucleosome positioning within individual rDNA repeats. Previous experiments analyzing the chromatin structure of the rDNA have used micrococcal nuclease (Fritze *et al.*, 1997; Cioci *et al.*, 2002; Li *et al.*, 2006a) or psoralen (Smith and Boeke, 1997) to show that in strains lacking Sir2 there is a change in chromatin accessibility. In other words, in cells lacking Sir2 chromatin forms a more open structure that allows increased accessibility to MNase or psoralen. Sir2 affects silencing at the NTS region more than at the transcribed regions and the loss of Sir2 increases the number of actively transcribed genes at the rDNA (Smith and Boeke, 1997). Studies focusing specifically on NTS2 show the positioning of five nucleosomes changed in cells lacking Sir2 that result in increased

MNase accessibility. In these strains, increased MNase accessibility is found at the areas between specific nucleosomes (Cioci *et al.*, 2002) at linker regions downstream of nucleosome 1, between nucleosome 2 and 3, and upstream of nucleosome 5 (Li *et al.*, 2006a) (Figure 2b).

Hmo1 is part of the high-mobility group B (HMGB) protein family and contains a DNA-binding motif. Hmo1 associates with the active transcribed 35S rRNA genes. The discovery of Hmo1 is an important one because of its similarity to human UBF, a homolog of yeast UAF (Schnapp *et al.*, 1994). In many ways human UBF and yeast UAF are similar however yeast UAF lacks the conserved HMG-box characteristic of human UBF. A HMG-box domain allows binding to DNA and is a characteristic of HMG proteins (Stros *et al.*, 2007). UAF lacks this domain however there is no question it associates within the promoter region of RNA Pol I. Hmo1 contains such a domain and may allow direct binding of RNA Pol I to chromatin. Hmo1 associates throughout the entire actively transcribed 35S rRNA gene and is especially enriched at the RNA Pol I promoter region. A technique called ChEC for Chromatin Endogenous Cleavage was used to study chromatin structure *in vivo* (Merz *et al.*, 2008). Using this technique, a protein of interest (i.e. Hmo1) is fused to MNase and expressed *in vivo* to determine where the protein of interest associates with chromatin. By analyzing cut sites in genomic DNA generated from this technique, the chromatin present at the promoter region of RNA Pol I was found to loop around UAF. MNase connected to histones in a yeast strain with an all actively transcribed array suffered less cleavage in the rDNA (and thus has fewer histones) than an array containing 50% active and 50% inactive 35S

rRNA genes. ChEC and psoralen can be combined to show that Hmo1 associates with transcriptionally active genes. Hmo1 is specific to active repeats only and remains bound to the RNA Pol I promoter region once RNA Pol I leaves the promoter region and enters elongation (Merz *et al.*, 2008). Remember that TBP, CF, and UAF also remain bound to the promoter region when RNA Pol I enters elongation (Aprikian *et al.*, 2001; Bier *et al.*, 2004; Claypool *et al.*, 2004). To summarize, the promoter region of actively transcribed genes contain TBP, CF, UAF, and Hmo1.

The data generated above from MNase and psoralen experiments about active and inactive repeats are based on a population of rDNA repeats in a population of cells. Research has been branching out into single molecule and single cell experiments. By using fluorescence *in situ* hybridization (FISH) in individual yeast cells, data was collected on the kinetics of rDNA transcription by RNA Pol I at the 35S rRNA gene (Tan and van Oudenaarden, 2010). Using FISH, the rate of the 35S rRNA gene switching from inactive to actively transcribed by RNA Pol I, or activation rate, and the average number of transcripts produced from a single activation event, or 'burst size' were measured. By measuring RNA expression from single rDNA repeats over time from a single cell the activation rate is ~5 minutes before the 35S rRNA gene becomes inactive again (Tan and van Oudenaarden, 2010). For each active repeat, approximately 190 transcripts are produced during those 5 minutes (Tan and van Oudenaarden, 2010). These results shows that reinitiation of RNA Pol I takes ~1.6 seconds, which agrees with previous data from electron microscopy experiments that estimated between 1.2 and 2.2 seconds (French *et al.*, 2003). Data from these experiments support the idea that at any

one time 50% of repeats are active and 50% are inactive (Dammann *et al.*, 1993) and that the state of transcription within individual repeats is not inherited between cell cycles (Lucchini and Sogo, 1995). Using FISH researchers were able to go a step further to analyze a single cell with a single rDNA repeat. They propose through mathematical modeling that throughout an entire cell cycle all repeats will eventually become actively transcribed by RNA Pol I with the proportion of active and inactive repeats (50/50) staying constant at any one time (Tan and van Oudenaarden, 2010). These results would be impossible to replicate with our single molecule technique with *M.CviPI*. Our experiments with *M.CviPI* were based on the idea that the proportion of active repeats within a cell remain active throughout the entire cell cycle and that inactive repeats remain silent through the entire cell cycle. As mentioned above, there is only a 5 minute window when the 35S rRNA is actively transcribed and accordingly the nucleosome positioning within NTS2 is in an open conformation. Our induction times of 60 and 120 minutes for *M.CviPI* are much too long to properly characterize a 35S rRNA gene that is only actively transcribed for 5 minutes, but these induction times are necessary for this technique. This is discussed in greater detail in Chapter II.

Regulator of nucleolar silencing and telophase exit

The rDNA forms a distinct structure within the nucleus called the nucleolus that is positioned on one side of the periphery of the nucleus forming a dark nuclear cap structure (Santoro, 2005). Nucleolar structure can change from a compact cap seen on the edge of the nucleus to an amorphous blob due to factors that damage nucleolar structure. Silencing of RNA Pol II transcription at the rDNA occurs through the RENT

complex for Regulator of Nucleolar silencing and Telophase exit and is comprised of Sir2, Net1, and Cdc14 (Shou *et al.*, 1999; Straight *et al.*, 1999). RENT is recruited to two regions of the rDNA: by Fob1 in NTS1 and by RNA Pol I at the 35S promoter region in NTS2 (Huang and Moazed, 2003). Net1 (Nucleolar silencing Establishing factor and Telophase regulator) is known to promote RNA Pol I mediated transcription at the rDNA and is believed to tether RNA Pol I specifically to the nucleolus by a physical interaction with RENT (Shou *et al.*, 2001; Huang and Moazed, 2003). Data has shown that deletion of Net1 leads to loss of a specific subunit of RNA Pol I causing decrease in rDNA transcription (Shou *et al.*, 2001). Net1 is specifically required for silencing at the rDNA and recruits Sir2 (Straight *et al.*, 1999) and Cdc14 (Shou *et al.*, 1999) to the nucleolus silencing transcription. Deletion of Net1 causes a loss of gene silencing at the rDNA and instability in nucleolar structure resulting in the distinct cap-like structure of the nucleolus to disperse (Straight *et al.*, 1999; Shou *et al.*, 2001).

In addition to silencing of RNA Pol II transcription at the rDNA, RENT also plays a role in exit from mitosis. Exit from mitosis in yeast requires two signaling cascades: Cdc fourteen early anaphase release (FEAR) and mitotic exit network (MEN). FEAR functions to release Cdc14 during early stages of anaphase which stimulates MEN to maintain release of Cdc14 during later stages of anaphase leading to exit from mitosis (Stegmeier *et al.*, 2002). Net1's recruitment of Cdc14 to the nucleolus inhibits Cdc14 activity and results in inhibiting exit from mitosis (Shou *et al.*, 1999; Traverso *et al.*, 2001), however once Net1 is phosphorylated at the beginning of anaphase by the protein kinase Cdk1, Cdc14 (Azzam *et al.*, 2004) and Sir2 are released, while Net1

remains associated at the rDNA throughout the cell cycle. The release of Cdc14 allows phosphorylation of its substrates and exit from mitosis (Shou *et al.*, 1999; Straight *et al.*, 1999).

RENT also plays a role in repressing recombination between repetitive rDNA repeats. There are two key proteins that control recombination at the rDNA array in yeast: Sir2 (Gottlieb and Esposito, 1989) and Fob1 (Kobayashi and Horiuchi, 1996). As mentioned above, RENT binds within NTS1 and the 35S promoter region of RNA Pol I in NTS2 (Huang and Moazed, 2003). While bound to NTS1, RENT associates with Fob1 at the RFB site. This association of RENT to Fob1 is required for silencing of RNA Pol II transcription in NTS1 (Huang and Moazed, 2003). Fob1 is known to enhance recombination but its activity is repressed through its recruitment of RENT to NTS1 (Kobayashi and Horiuchi, 1996; Huang and Moazed, 2003). With Fob1 bound at the RFB site (Kobayashi *et al.*, 1992), replication forks traveling left from the ARS in NTS2 are blocked from colliding with RNA Pol I transcriptional machinery (Kobayashi and Horiuchi, 1996; Kobayashi, 2003). Thus, NTS1 is a region that has a role in repressing recombination and in halting replication forks (Kobayashi and Horiuchi, 1996; Kobayashi *et al.*, 1998).

In cells lacking Sir2, recombination increases between 10-15% and this increase is specific to the rDNA (Gottlieb and Esposito, 1989). Sir2 represses unequal sister-chromatid recombination occurring in NTS1 (Kobayashi *et al.*, 2004). With a loss of Sir2, two paired alleles of the rDNA recombine between each other leading to the two copies of the rDNA, the sister chromatids, to misalign and the copy number of the rDNA

to increase in one daughter cell and to decrease in the other daughter cell. When a double strand break occurs between sister chromatids, Sir2 uses an association with cohesin to repress recombination between sister chromatids during mitosis (Kobayashi *et al.*, 2004). Cohesin is a multi-subunit ring-like complex that associates within NTS1 at the replication fork to keep sister chromatids together by encircling the DNA until the cell reaches anaphase and chromatids separate (Lengronne *et al.*, 2006). Cohesin is cleaved by separase once sister chromatids have properly aligned in metaphase. Cleavage of cohesin leads to anaphase and segregation of sister chromatids (Haering *et al.*, 2008).

Cohesin and condensin at the rDNA

In 2006 Huang *et al.* proposed the existence of a "protein bridge" containing RENT, Fob1, and other proteins that span between NTS1 regions and functions to repress recombination at the rDNA between sister chromatids during replication (Huang *et al.*, 2006a). By purifying proteins known to function in rDNA silencing (Sir2, Net1, and Fob1) a series of proteins were identified that included nuclear envelope proteins, Tof2, Lrs4, and Csm1. Tof2 was originally discovered to co-purify with the RENT complex in a study focusing on identifying additional Sir2 interacting factors (Tanny *et al.*, 2004). Tof2 and a heterodimer of Lrs4 and Csm1 specifically associate with Fob1 and RENT at NTS1 to silence an RNA Pol II gene artificially inserted into NTS1. The Huang *et al.* 2006 article in combination with other articles that detail the association of Lrs4 and Csm1 with cohesin through screens (Graumann *et al.*, 2004) and one that showed the association of Sir2 with cohesin at the rDNA in NTS1 (Kobayashi *et al.*,

2004; Kobayashi and Ganley, 2005) proposed a model that cohesin forms a ring specifically at NTS1 and associates with RENT, Fob1, Tof2, Lrs4, and Csm1 to clamp sister chromatids together. The purpose of this clamp is to prevent recombination between sister chromatids that could result in a change in the number of repeats present at the rDNA array.

An important step to cell division is the ability to properly compact and segregate the ~1-2 Mb rDNA array during mitosis. During anaphase of mitosis, many steps are involved in controlling segregation of chromosomes. Tof2, Lrs4, Csm1, Fob1 and Cdc14 all function in segregation of the rDNA during the cell cycle, and associate with condensin, another large ring complex of proteins that functions to compact chromosomes for segregation during mitosis (Geil *et al.*, 2008; Waples *et al.*, 2009). Both cohesin (Graumann *et al.*, 2004; Kobayashi *et al.*, 2004; Kobayashi and Ganley, 2005) and condensin (Johzuka *et al.*, 2006; Machin *et al.*, 2006) have been shown to associate at NTS1 in the rDNA. During anaphase, RNA Pol I transcription is inhibited by Cdc14 by preventing the association of Pol I with the rDNA. This inhibition of transcription is required for association of condensin (Clemente-Blanco *et al.*, 2009). Deletion of Tof2, Lrs4, Csm1, or Fob1 abolishes the compacted structure of the rDNA created by association of condensin and therefore affects segregation during the cell cycle, suggesting a role of these four proteins in recruitment of condensin and compaction of the rDNA (Geil *et al.*, 2008; Waples *et al.*, 2009).

A model can be proposed to explain how the association of condensin and cohesin is coordinated at the rDNA (Figure 4). At NTS1, RENT (Sir2, Net1, and

Cdc14) interacts with Tof2 and Fob1. Prior to anaphase, cohesin functions to restrict recombination at the rDNA through the formation of a protein bridge formed around sister chromatids as discussed above (Huang *et al.*, 2006a; Haering *et al.*, 2008). Cohesin is lost by the time the cell reaches anaphase due to cleavage by separase. At the beginning of anaphase, Net1 is phosphorylated by Cdk1 and a partial release of Cdc14 occurs (Azzam *et al.*, 2004). This partial release of Cdc14 inhibits RNA Pol I transcription (Clemente-Blanco *et al.*, 2009) and causes a late association of condensin to the rDNA (Sullivan *et al.*, 2004; Geil *et al.*, 2008; Waples *et al.*, 2009). Condensin is known to associate with chromatin throughout the cell cycle at relatively even levels (D'Ambrosio *et al.*, 2008), but once there is a partial release of Cdc14 a portion of condensin is siphoned off and relocated to the rDNA. This early release of Cdc14 by FEAR not only initiates exit from mitosis but also leads to chromosome segregation (D'Amours *et al.*, 2004). The remaining Cdc14 is retained by Tof2, either directly or by preventing full release of Cdc14 from Net1. This retention of Cdc14 at the rDNA is required for the rDNA to compact and segregate in late anaphase. Once late anaphase is reached (MEN pathway) the rest of Cdc14 is released, chromosomes segregate and cells exit mitosis to begin the cell cycle again (Shaw and Doonan, 2005; Waples *et al.*, 2009). Thus, the release and retention of Cdc14 controls exit from mitosis and chromosome segregation.

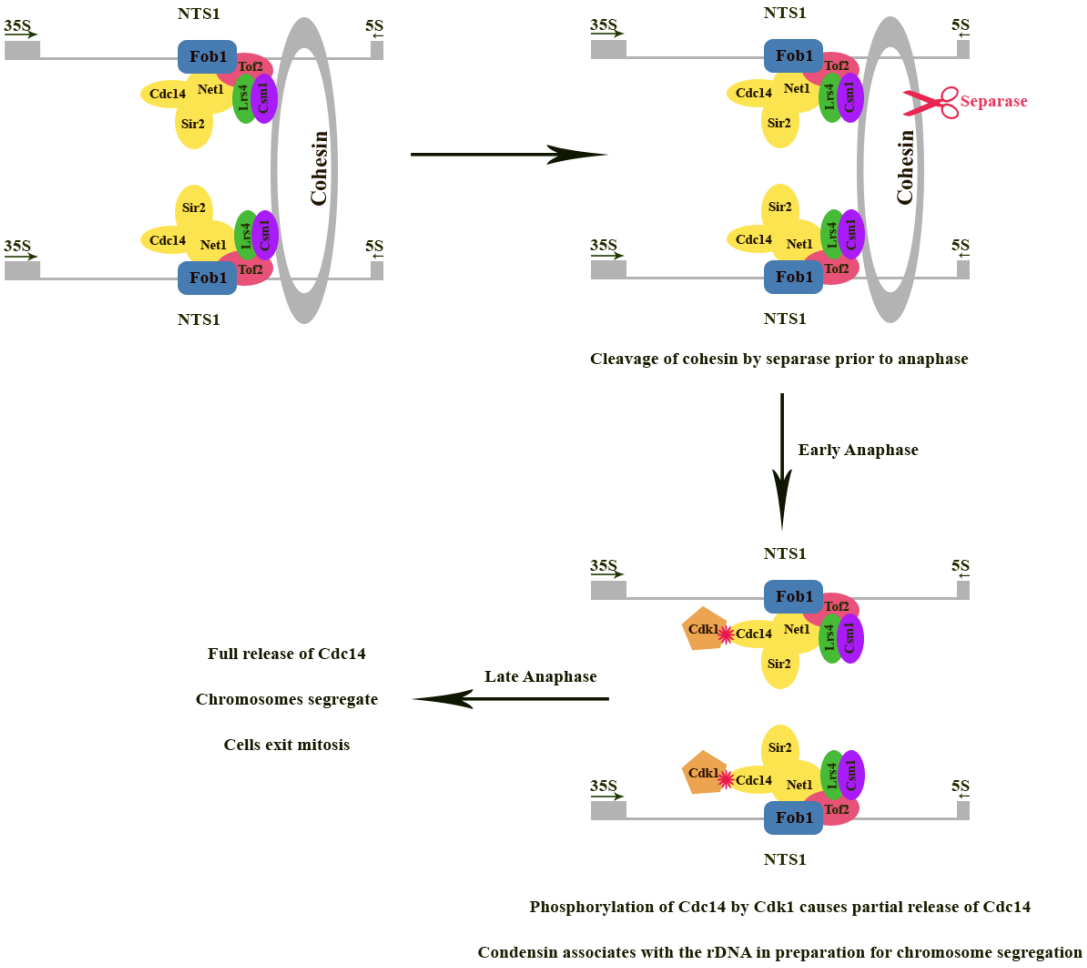


Figure 4. Cohesin and Condensin at the rDNA. Within NTS1, RENT (a heterotrimer in yellow) associates with Fob1 (blue) and Lrs4 (green), Csm1 (purple), and Tof2 (pink). These proteins associate with cohesin as shown in the image. Association of these proteins represses recombination at the rDNA between sister chromatids during replication. A partial release of Cdc14 marks early anaphase leading to association of condensin in preparation for chromosome segregation. Late anaphase requires the full release of Cdc14 and leads to chromosome segregation and exit from mitosis.

Chromosome linkage inner nuclear membrane proteins

As with the other heterochromatin regions in yeast, the rDNA is known to associate with the nuclear periphery. The rDNA forms a distinct structure called the nucleolus. During the purification of Lrs4 and Csm1 by Huang *et al.* 2006 nuclear proteins were found to associate with the protein clamp described above. Further research showed that the "protein bridge" of Lrs4 and Csm1 associates with the nuclear periphery through chromosome linkage inner nuclear membrane proteins (CLIP) composed of Heh1 and Nur1 (Mekhail *et al.*, 2008). CLIP is required for repressing recombination between rDNA repeats, but on its own CLIP is not required for silencing of RNA Pol II genes at the rDNA. Localization to the nuclear periphery is required to repress recombination since losing either CLIP or Lrs4/Csm1 causes the structure of the nucleolus to destabilize and an increase in recombination. Together, the association of CLIP with RENT and Lrs4/Csm1 at NTS1 functions to tether the rDNA to the nuclear periphery, stabilizes nucleolar structure, and prevents recombination between rDNA repeats (Mekhail *et al.*, 2008).

In human and mouse cells, recruitment to the nuclear periphery alters gene expression by silencing transcription of an artificially inserted gene and HDACs play some role in silencing transcription (Finlan *et al.*, 2008; Reddy *et al.*, 2008). Early research in human and mouse model systems showed that euchromatin and heterochromatin regions localize to the nucleoplasm and nuclear periphery, respectively (Ferreira and Carmo-Fonseca, 1997; Ferreira *et al.*, 1997; Sadoni *et al.*, 1999). Histones located near the nuclear periphery are hypoacetylated and those located away from the

periphery are hyperacetylated (Sadoni *et al.*, 1999). In other words, localization to the nuclear periphery is associated with silenced transcription and contains histone modifications characteristic of silent chromatin (Reddy *et al.*, 2008). Although a correlation has been shown between association with the nuclear periphery, H4 deacetylation, and decreased transcription (Sadoni *et al.*, 1999; Finlan *et al.*, 2008; Reddy *et al.*, 2008), evidence pointing to a specific HDAC has not been found (Ruault *et al.*, 2008).

Summary

Heterochromatin is described as a closed chromatin structure silent for transcription. The previous description of telomeres in *S. cerevisiae* provides a classic example of heterochromatin. The rDNA contradicts the classic definition of heterochromatin because it contains high levels of transcription by RNA Pol I and RNA Pol III. Silencing of transcription at the rDNA is specific to RNA Pol II transcription and unlike telomeres requires only Sir2 of the SIR family proteins. The rDNA is further complicated by results showing that only half of the 35S rRNA genes are actively transcribed in five minute intervals while to other half are silent. This fast switching between active and inactive 35S rRNA genes is discussed in greater detail in Chapter II. Hmo1 has been shown using a technique called ChEC to specifically associate with actively transcribed genes only. Our data shown in Chapter II might represent the association of Hmo1 in the rDNA.

STRUCTURE AND MECHANISM OF HUMAN SET DOMAIN PROTEINS

Post-translational modifications of histones is a field that encompasses many types of modifications, most commonly histone methylation, acetylation and ubiquitylation. Histone methylation is a modification that has been linked to both euchromatin and heterochromatin regions. In Saccharomyces cerevisiae, H3K4 methylation is associated with high levels of histone methylation in euchromatin regions and low levels in heterochromatin regions. By studying the crystal structures of Set7/9 and MLL1, we can further our understanding of lysine methylation and pinpoint important residues in SET domain proteins. Only two human H3K4 methyltransferases have been crystallized: Set7/9 and MLL1. This discussion will focus on Set7/9 and MLL1, and later on in Chapter III their relevance to yeast Set1. To help understand these comparisons, Table 2 shows the conserved amino acids discussed among these three proteins as well as mutants made in yeast Set1 discussed in Chapter III. SET domain methyltransferases exhibit many similarities: the presence of a pseudoknot, conserved tyrosines and phenylalanines, and a water channel to deprotonate lysine 4 of histone H3. SET domain proteins contain conserved motifs of GXG (I), YXG (II), RXXNHS/C (III), EELXY/FXY (IV) and CXCXXXXC (V) (Figure 5). Each motif has a specific role in SET domain proteins. GXG (I) motifs are known for their role in binding the cofactor S-adenosyl methionine in methyltransferase enzymes (Shields *et al.*, 2003). RXXNHS/C (III) and EELXY/FXY (IV) along with YXG (II) form the active site within a substrate specific channel for the target lysine residue and interact with lysine 4 of histone H3 and SAM (Kwon *et al.*, 2003; Xiao *et al.*, 2003). CXCXXXXC

(V) coordinates a zinc ion that aides SET domain proteins in properly folding the c-terminus (Zhang *et al.*, 2003; Southall *et al.*, 2009). Those that lack these conserved cysteines rely on a c-terminal α helix to properly associate with the larger protein structure and form the lysine substrate specific channel.

Table 2. Comparison of conserved amino acids between yeast and human SET domain proteins.

Set1-COMPASS	**Set1 mutants**	**Set7/9**	**MLL1**
Saccharomyces cerevisiae		**Human**	**Human**
G951	G951A	G229	G3842
Y967	Y967F/A	Y245	Y3858
I972	I972A	I250	I3863
I989	I989A	L267	I3880
Y993	Y993A	E271	Y3883
R1013	R1013H	H293	R3903
N1016	N1016A	N296	N3906
H1017	H1017A/R/L	H297	H3907
T1024	T1024Y/A	Y305	Y3914
R1042	R1042H/A	R323	R3932
E1048	E1048A	E329	E3938
E1049	E1049A/G	E330	E3939
L1050	L1050A	L331	L3940
Y1052	Y1052F/A	V333	Y3942
Y1054	Y1054F/A	Y335	Y3944
F1056	F1056Y/A	Y337	F3946
C1068	C1068A	--	C3957
C1070	C1070A	--	C3959
C1075	C1075A/Y	--	C3964

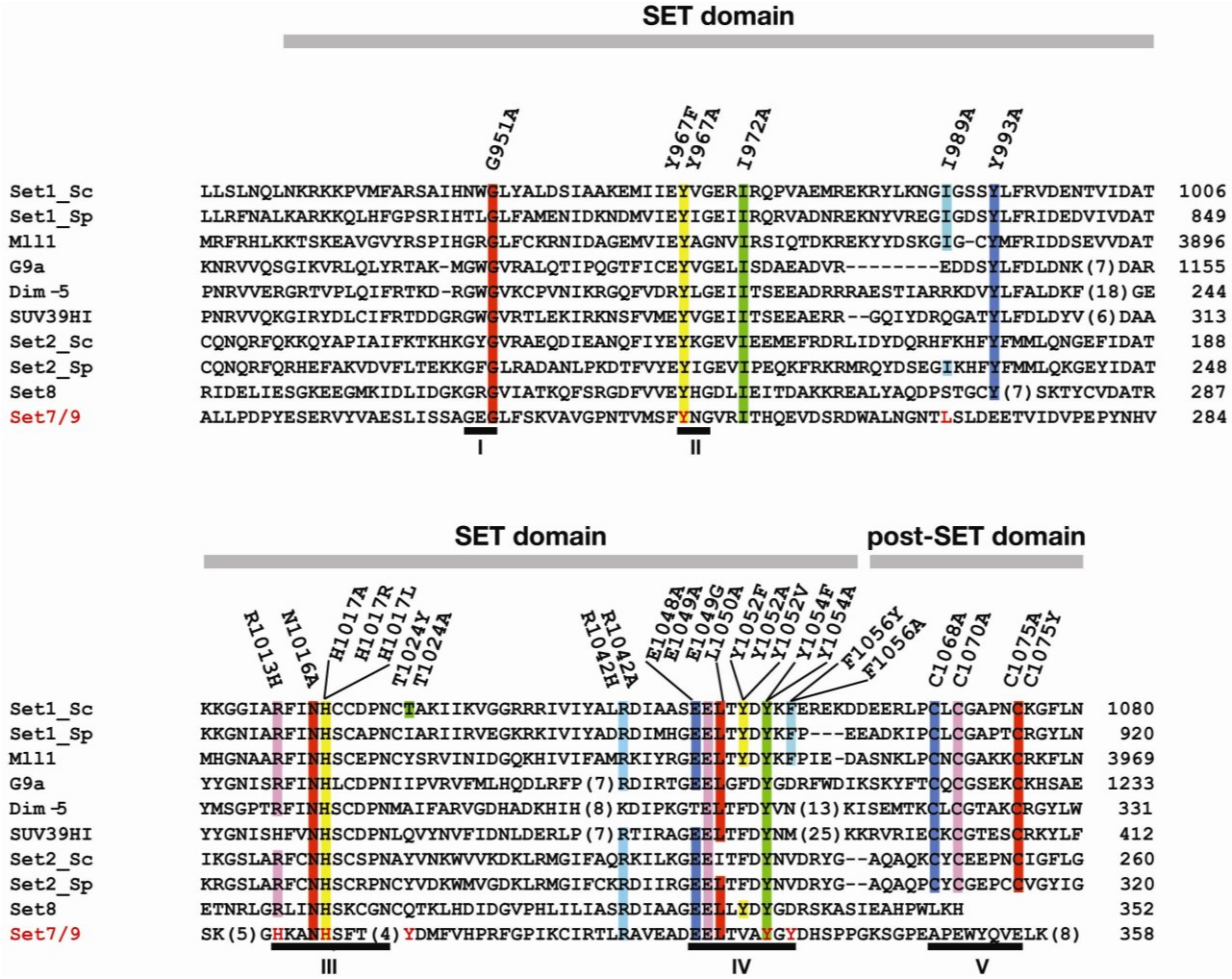


Figure 5. Alignment of SET domain proteins. Alignment of SET domain proteins from several eukaryotes. Conserved motifs are signified by I-V. Conserved amino acids are highlighted with the substitutions in yeast Set1 (discussed in Chapter III) indicated above the alignment. The abbreviations Sc and Sp refer to *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, respectively.

SET domain proteins contain a unique structure called a pseudoknot formed from one strand threading through the loop of another. This knotted structure is characteristic of SET domain proteins and brings into contact two conserved motifs RXXNHS/C (III) and EELXY/FXY (IV) that together form the active site for the majority of all SET domain proteins. This structure was first identified in Set7/9 but can also be seen in MLL1 as well (Jacobs *et al.*, 2002) (Figure 6). Structural studies of Set7/9 and MLL1 have provided information on key amino acids and their role in interacting with substrates and their role in the methylation mechanism. Set7/9 has been extensively studied through crystal structures combined with theoretical studies utilizing computer simulations to gain information on the methyltransferase mechanism at the active site. Computer simulations allowed researchers to use a previously crystallized structure of Set7/9 (Xiao *et al.*, 2003) to simulate certain amino acid mutations and the role of water in the catalytic mechanism (Hu and Zhang, 2006; Wang *et al.*, 2007b; Zhang and Bruice, 2007; Hu *et al.*, 2008). From these crystal structures and computer simulations, general statements can be made about important amino acids listed in Table 2.

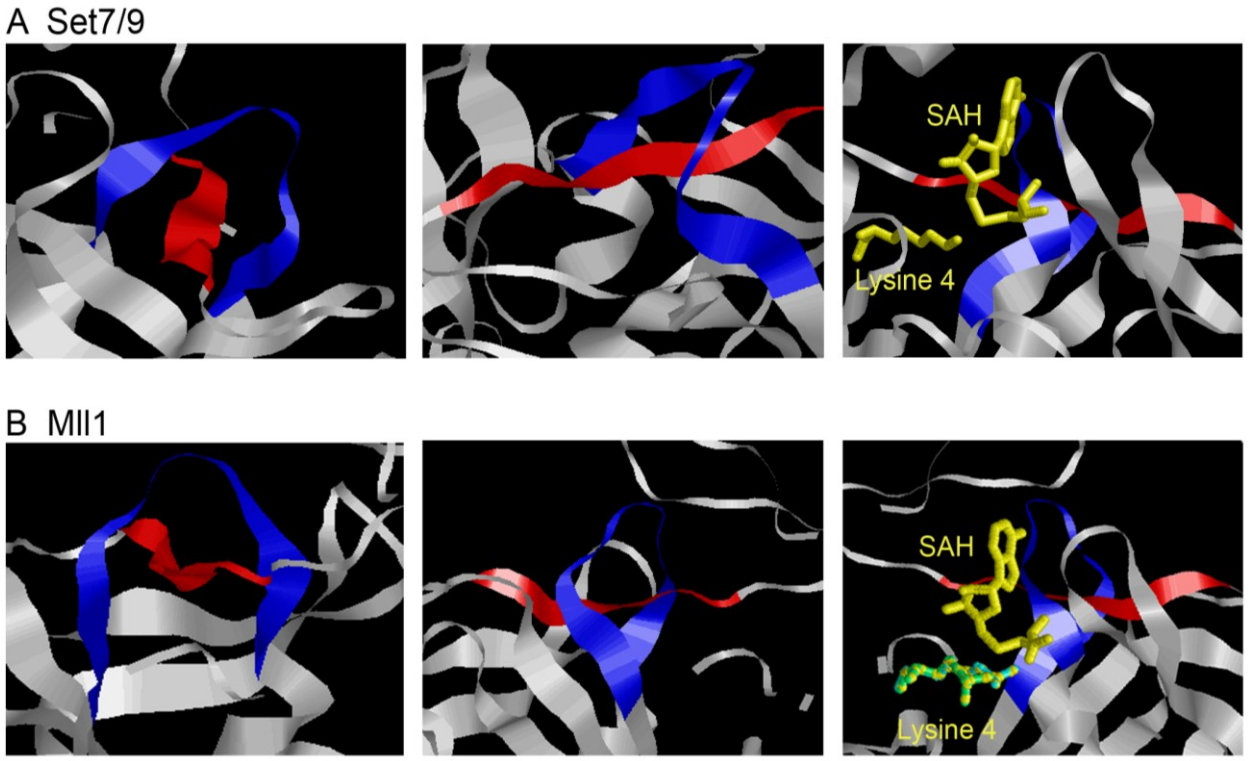


Figure 6. High resolution images of the pseudoknot structure within human Set7/9 and MLL1. **(A)** The Set7/9 crystal structure was generated by Xiao *et al.* 2003 and visualized using Rasmol version 2.7.5 (Sayle and Milner-White, 1995; Bernstein, 2000). Each of the three panels is a ribbon image of Set7/9 pseudoknot from different angles, with the third containing S-adenosyl homocysteine (SAH) and lysine 4 of histone H3 in yellow. Amino acids 295-305 are highlighted in blue and represent the RXXNHS/C motif. Amino acids 329-335 are highlighted in red and represent the EELXY/FXY motif. **(B)** The MLL1 crystal structure was generated by Southall *et al.* 2009 and visualized using Rasmol version 2.7.5. Again, each of the three panels is a ribbon image of MLL1 pseudoknot from different angles; however, the third panel shows S-adenosyl homocysteine (SAH) in yellow with dimethylated lysine 4 in yellow and green in the two different positions it maintains when bound. Amino acids 3905-3915 are highlighted in blue and represent the RXXNHS/C motif. Amino acids 3938-3944 are highlighted in red and represent the EELXY/FXY motif.

Tyrosine residues control active site structure and product specificity

Set7/9 is characterized by a substrate specific channel that runs through the protein, with S-adenosyl methionine (SAM) and lysine 4 of histone H3 (H3K4) associating through opposite sides of the protein through this channel (Kwon *et al.*, 2003; Xiao *et al.*, 2003). This substrate specific channel is lined with conserved amino acids that make contact with both SAM and lysine 4 and form the structure of the channel: Tyr245, Leu267, Tyr305, Tyr335, and Tyr337. The hydroxyl side chains of Tyr245, Tyr305, and Tyr335 are directed toward the methyl group of SAM and lysine 4 of histone H3 (Kwon *et al.*, 2003; Xiao *et al.*, 2003). Figure 7 contains high resolution images of the crystal structure of Set7/9 showing the core amino acids that form the structure of the substrate specific channel (Figure 7a) along with other conserved amino acids present in motifs within SET domain proteins (Figure 7b,c). Of the tyrosine residues listed Tyr335 is a key residue that forms the active site and its positioning is stabilized by His297. His293 and His297 are conserved residues that serve a structural role. In the crystal structure of Set7/9 both histidines are positioned away from the active site making it unlikely that these residues serve any catalytic role. His293 and His297 may assist Tyr245 and Tyr335, respectively, with catalysis by properly positioning these tyrosine residues in the active site (Kwon *et al.*, 2003; Xiao *et al.*, 2003).

The majority of what is known about the catalytic mechanism of Set7/9 comes from theoretical studies using computer simulations. These simulations have focused on studying Tyr245 and Tyr305, and have concentrated on the formation of a water channel

important in deprotonation of the amine group of lysine 4 of histone H3. In SET domain proteins there is a phenomenon that exists called a phenylalanine/tyrosine switch that controls product specificity. The presence of the hydroxyl group of a tyrosine residue blocks the rotation of methylated lysine 4 of histone H3 and prevents higher levels of methylation (Collins *et al.*, 2005). SET domain proteins are rich in tyrosine residues, any one of which could control product specificity. Set7/9 is a human H3K4 methyltransferase that is only able to add a single methyl group to lysine 4 (Wilson *et al.*, 2002; Kwon *et al.*, 2003; Xiao *et al.*, 2003). In Set7/9, Tyr245 and Tyr305 are believed to serve this function, while in other SET domain proteins other tyrosine residues may fulfill this role. These two amino acids were studied by artificially simulating Y245A and Y305F mutants and studying the effect of these mutants through computer modeling (Hu and Zhang, 2006; Zhang and Bruice, 2007; Hu *et al.*, 2008). These mutants were also studied by crystallizing Set7/9 with these mutations along with an alternative substrate TAF10 peptide to further analyze their effect on histone methylation (Del Rizzo *et al.*, 2010). These results are discussed in Appendix C.

Through these computer simulations Set7/9 is a mono methyltransferase only, due to steric hindrance by Tyr245. Tyr245 interacts with lysine 4 of histone H3 through its ring. Removing this ring by simulating the mutant Y245A through computer modeling removes steric hindrance within the active site and allows Set7/9 to trimethylate H3K4 where normally it only monomethylates (Xiao *et al.*, 2003; Hu and Zhang, 2006). This mutation formed the model that the presence of a tyrosine can hinder higher-ordered methylation (Xiao *et al.*, 2003). Structurally, the presence of the

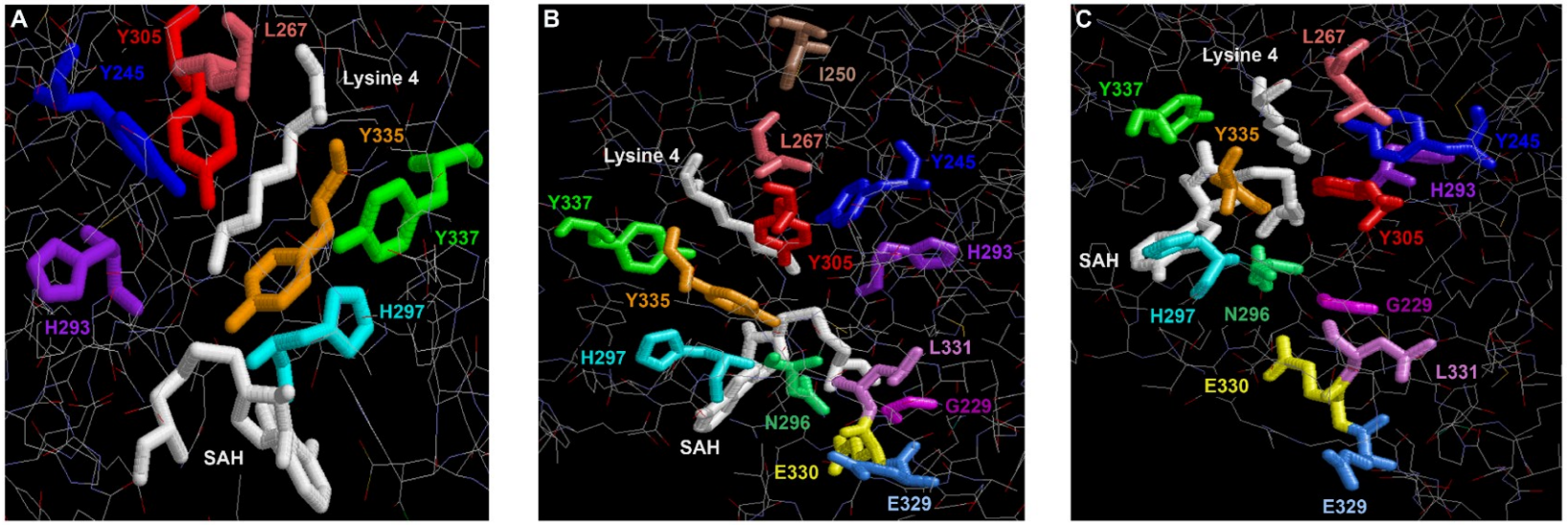


Figure 7. High resolution images of conserved amino acids in the crystal structure of human Set7/9. The Set7/9 crystal structure was generated by Xiao *et al.* 2003 and visualized using Rasmol version 2.7.5 (Sayle and Milner-White, 1995; Bernstein, 2000). The color of amino acids is consistent between all images with S-adenosyl homocysteine and lysine 4 of histone H3 in white. **(A)** Core conserved amino acids as described within the text. **(B,C)** Different views of Set7/9 with conserved amino acids, S-adenosyl homocysteine (SAH) and lysine 4 of histone H3.

hydroxyl group on Tyr245 prevents monomethylated H3K4 from rotating and aligning with a new SAM molecule for further methylation (Hu and Zhang, 2006). This misalignment of the amine group of lysine 4 causes a decrease in the amount of product produced over time (Del Rizzo *et al.*, 2010).

Tyr305 is also studied for its role in product specificity. The active site contains several water molecules, one of which is close to the amine group of lysine 4 of histone H3 and forms a hydrogen bond with Tyr305 (Kwon *et al.*, 2003; Xiao *et al.*, 2003; Wang *et al.*, 2007b). The mutation Y305F causes Set7/9 to mono- and dimethylate H3K4 (Hu *et al.*, 2008). With Y305F, the loss of the hydroxyl group disrupts hydrogen bonds within the active site, specifically with a single water molecule (Kwon *et al.*, 2003). Within the active site, hydrogen bonds formed between monomethylated H3K4, Tyr245, and Tyr305 work together to prevent rotation of the methylated H3K4 and prevent further methylation (Zhang and Bruice, 2007).

Formation of a water channel is critical for the deprotonation of lysine 4

Hydrogen bonding between amino acids within the active site of Set7/9 can align both substrates lysine 4 of histone H3 and SAM properly for monomethylation to occur, along with positioning methylated lysine 4 into a suitable position for higher levels of methylation (Couture *et al.*, 2006c). As mentioned above, further methylation by hydrogen bonds forming between Tyr245 and Tyr305 with monomethylated lysine 4 can prevent the alignment of SAM for further methylation (Zhang and Bruice, 2007). The role of hydrogen bonding within the active site became increasingly important by the discovery of a string of water molecules that hydrogen bond within the active site,

termed a water channel (Zhang and Bruice, 2007). The formation of this water channel is critical for the deprotonation of the amine group of lysine 4 of histone H3 to occur, the first step in the mechanism for methylation of H3K4. Through computer simulations with Set7/9, Y245A and Y305F allow an additional water channel to form following monomethylation of H3K4 by disrupting the structure of the active site and/or by disrupting the strict hydrogen bonds formed within the active site. This additional water channel allows further methylation of H3K4 to occur (Zhang and Bruice, 2007; Hu *et al.*, 2008; Zhang and Bruice, 2008a).

The mechanism of Set7/9 relies on the formation of a water channel to deprotonate the amine group of lysine 4 of histone H3 through an SN2 reaction (Dirk *et al.*, 2007; Zhang and Bruice, 2007) (Figure 8). The water channel functions as a proton acceptor from the amine group of lysine 4 to allow deprotonation of lysine 4. A conflicting model has been proposed that Tyr335 acts as a base to deprotonate the amine group of lysine 4 of histone H3 (Guo and Guo, 2007). Using computer simulations, the pKa of amino acids within the active site can be calculated in different environments. When Set7/9 is bound with SAM, the pKa of the amine group of lysine 4 is 8.5, meaning that when both substrates are bound deprotonation can occur. However, the pKa of Tyr335 in Set7/9 with both bound substrates is 18.7, making it unlikely to function as a base to deprotonate lysine 4 (Zhang and Bruice, 2007). Once deprotonation occurs the reaction follows through to methylate H3K4. Binding of lysine 4 of histone H3 or SAM is either a random or ordered event, but once bound deprotonation of lysine 4 occurs followed by methylation.

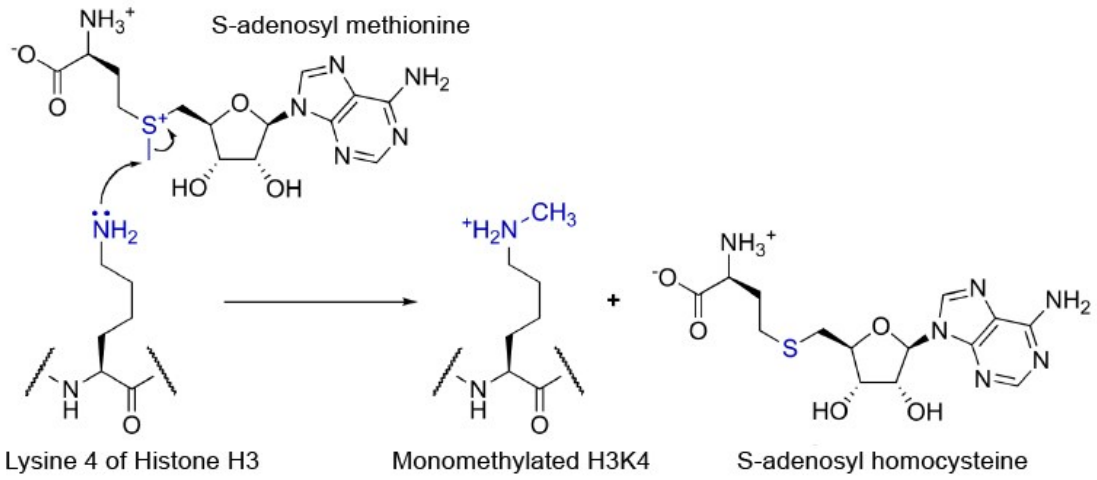


Figure 8. SN2 reaction catalyzed by SET domain proteins[2]. Water molecules within the active site function to deprotonate lysine 4 and allow a nucleophilic attack of a methyl group on S-adenosyl methionine.

[2] Reprinted with permission from "Chemical mechanisms of histone lysine and arginine modifications" by Brian C. Smith and John M. Denu, (2009), Biochim Biophys Acta, 1789, 45-57, Copyright 2009 by Elsevier.

The model for SET domain proteins is that monomethylation is distributive while di- and trimethylation are successive. Unmodified H3K4 is methylated and released as monomethylated H3K4, suggesting a distributive model shown through the accumulation of monomethylated H3K4 throughout the course of an assay. Di- and trimethylation are successive; there is no accumulation of an intermediate suggesting that monomethylated H3K4 is carried straight on to trimethylated H3K4 without being released (Dirk *et al.*, 2007).

MLL1 is a superior model for understanding yeast Set1

MLL1 has also been studied structurally and is a better model for understanding yeast COMPASS than Set7/9 for two reasons. First, MLL1 forms a complex with subunits homologous to COMPASS unlike Set7/9 that functions as a monomer. Secondly, MLL1 retains key residues found in yeast COMPASS that are not conserved in Set7/9 such as Tyr3883, Tyr3942, and three conserved cysteines within motif V CXCXXXXC (Table 2). An important point about the crystal structure of MLL1 is that it is crystallized on its own without additional complex members. Because of this, proper folding or positioning of amino acids must be taken with a degree of uncertainty. For instance, there are great differences in the positioning of key active site residues between Set7/9 (Figure 7) and MLL1 (Figure 9). Tyr305 in Set7/9 is conserved as Tyr3914 in MLL1, and these residues are positioned in complete opposite directions in the structures of these two methyltransferases. His293 is thought to help properly position Tyr245 in Set7/9 (Kwon *et al.*, 2003; Xiao *et al.*, 2003); however, in MLL1

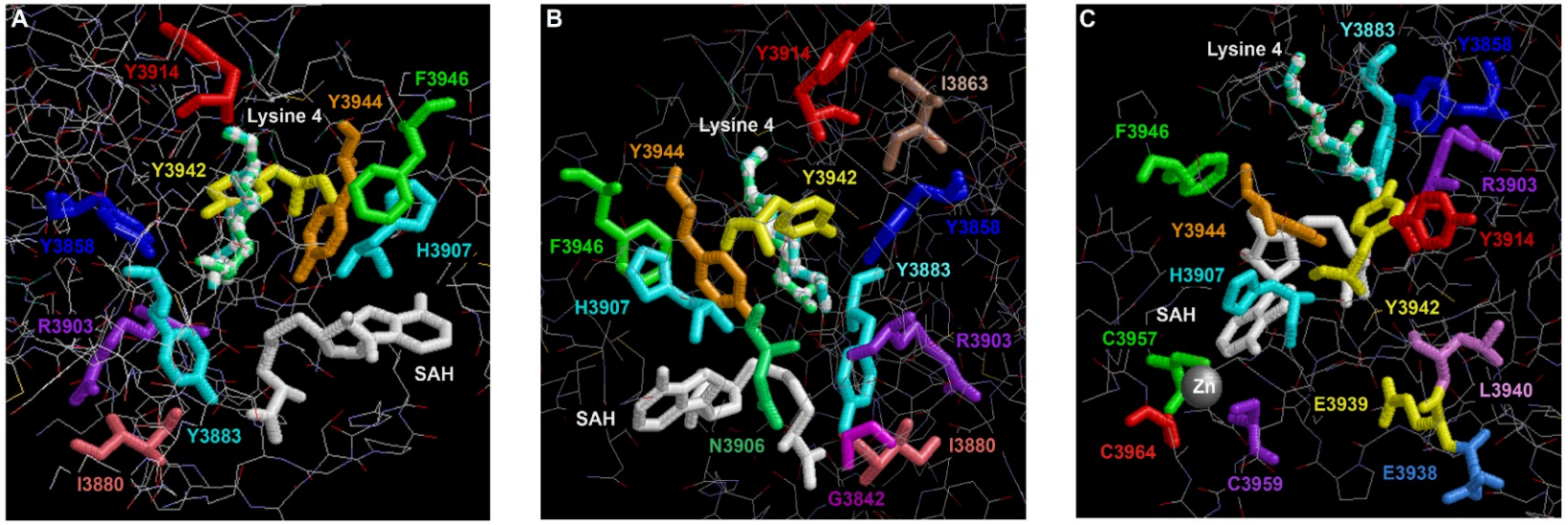


Figure 9. High resolution images of conserved amino acids in the crystal structure of human MLL1. The MLL1 crystal structure was generated by Southall *et al.* 2009 and visualized using Rasmol version 2.7.5 (Sayle and Milner-White, 1995; Bernstein, 2000). The color of amino acids is consistent between all images with S-adenosyl homocysteine (SAH) and dimethylated lysine 4 of histone H3 in white and cyan. **(A)** Core conserved amino acids as described within the text. **(B,C)** Different views of MLL1 with conserved amino acids, S-adenosyl methionine and dimethylated lysine 4 of histone H3.

Tyr3858 does not seem to be in close proximity to Arg3903 suggesting that this arginine serves no role in positioning Tyr3858.

The crystal structure of Set7/9 (Figure 7) shows a closed, substrate specific channel. However, the crystal structure of MLL1 shows an open channel that allows flexibility in the positioning of the histone H3 tail (Figure 9), but this open channel is thought to close upon binding of additional MLL1 complex members and shift the histone H3 tail into a single position that is found in other SET domain proteins (Southall *et al.*, 2009). The c-terminus of MLL1 contains motif V CXCXXXXC coordinates a zinc ion that assists in properly folding and forming a substrate specific channel for histone H3. The zinc ion is positioned by four cysteines; the three shown in Figure 9c along with a fourth cysteine Cys3909 within motif III RXXNHS/C that involves the c-terminus in the formation of the substrate specific channel (Southall *et al.*, 2009).

Without additional complex members bound, MLL1 alone is a monomethyltransferase that utilizes positioning of Tyr3942, which is not conserved in Set7/9, to prevent di- and trimethylation (Patel *et al.*, 2009). However, when bound with additional factors such as RBBP5, WDR5, and Ash2L, MLL1 is able to dimethylate H3K4. This contrasts with *in vivo* data that shows MLL1 mono-, di-, and trimethylates H3K4. Again, not all subunits of the MLL1 complex were included in these *in vitro* studies, and association additional factors may further alter the conformation of MLL1 and allow trimethylation (Patel *et al.*, 2009). In MLL1 Tyr3942 is considered the Phe/Tyr switch for this protein, and a Y3942F mutant in an *in vitro* system with MLL1,

RBBP5, WDR5, and Ash2L is able to trimethylate H3K4 (Patel *et al.*, 2009). The main active site resides are Phe3885, Tyr3942, Tyr3944, and Phe3946 (Figure 9). The substrate specific channel within MLL1 contains hydrogen bonds between Asn3906 and His3907 to the adenine ring of SAM and Tyr3883. Tyr3883 and Phe3885 are part of a YXF motif that has not been previously characterized. The orientation of Tyr3942 depends upon whether or not MLL1 binds a histone H3 peptide (Figure 9). Upon binding of a histone H3 peptide, the position of Tyr3942 orients and points towards Tyr3858; otherwise, Tyr3942 is positioned ~90° away pointing towards Tyr3914. The orientation of Tyr3942 shown in Figure 9 is the position it obtains when bound with a histone H3 peptide (Southall *et al.*, 2009).

DISSERTATION OVERVIEW AND SIGNIFICANCE

The scope of my research attempts to understand how histone methylation and chromatin structure affect gene silencing at the ribosomal DNA locus in *Saccharomyces cerevisiae*. Chapter II of this dissertation analyzes the chromatin structure of the NTS2 region within individual rDNA repeats by expressing a DNA methyltransferase *M.CviPI in vivo*, followed by bisulfite deamination, and cloning areas of interest. By expressing *M.CviPI in vivo*, we are able to study chromatin structure at the rDNA and visualize differences in chromatin structure between active and inactive repeats through differences in nucleosome positioning in NTS2. Actively transcribed repeats with an open chromatin structure would obtain high levels of methylation from *M.CviPI*, while inactive repeats with a closed chromatin structure would obtain low levels of

methylation. We based our experiments on the assumption that rDNA repeats that are actively transcribed remain in that state for the duration of a single cell cycle. Two approaches were taken in analyzing the chromatin structure of individual rDNA repeats. First, we focused on comparing wild type cells with cells lacking Sir2, searching for changes in nucleosome positioning similar to previously published results from out lab using MNase accessibility: downstream of nucleosome 1, between nucleosome 2 and 3, and upstream of nucleosome 5 (Li *et al.*, 2006a). Secondly, we attempted to differentiate repeats that are actively transcribed by RNA Pol I from repeats that are silent through nucleosome positioning within NTS2. Using this technique, we hope to create a new way of analyzing open and closed repeats at the rDNA and to better understand differences between euchromatin and heterochromatin.

Chapter III of this dissertation describes a mutagenesis of yeast Set1. Human H3K4 methyltransferases are a complex group in a complex system, with roles in many types of cancers and diseases. Studying Set1 in yeast provides a simpler system without sacrificing the relevance to higher eukaryotes. Mutants of Set1 were chosen by using a sequence alignment of other SET domain proteins, by studying the crystal structures of Set7/9, and through a random mutagenesis of the catalytic SET domain. Conserved tyrosine residues play a role in the majority of SET domain proteins, as well as motifs RXXNHS/C and EELXY/FXY that together form the active site of the majority of SET domain proteins. Mutants of Set1 were analyzed for their effect on silencing of a RNA Pol II-transcribed gene within the rDNA locus through northern analysis, histone methylation through quantitative Westerns, suppression of *ipl1-2* by plate assay,

association with COMPASS through TAP tag purifications, and an *in vitro* activity assay with full length H3 by Western analysis. This analysis of yeast Set1 mutants will increase our understanding of the importance of certain residues in the H3K4 methylation and cellular function of yeast Set1 and provide insight into Set1-like proteins in higher eukaryotes.

CHAPTER II

ANALYSIS OF CHROMATIN STRUCTURE AT NTS2 WITHIN INDIVIDUAL RDNA REPEATS USING *IN VIVO* EXPRESSION OF *M.CviPI*

The chromatin structure of the rDNA has been mapped by two *ex vivo* techniques: psoralen crosslinking and MNase accessibility. Both of these techniques provide results from a population of cells, each with a population of repeats. In order to understand the chromatin structure within individual repeats, the DNA methyltransferase *M.CviPI* was expressed *in vivo*, followed by bisulfite deamination, and cloning of the NTS2 region into plasmids. We have used this technique to study changes in nucleosome positioning as a result of a silencing defect caused by the loss of Sir2, and as an indicator of active transcription by RNA Pol I. Using this technique we can differentiate between open and closed chromatin structure by changes in nucleosome positions within NTS2 and by the presence of bound factors at the 35S rRNA gene promoter. From our results we see changes in nucleosome positions by comparing wild type and *sir2Δ* cells in S288c and W303 strain backgrounds, specifically in nucleosomes 5, 4, and 3. In addition, we used a 25 copy rDNA strain to compare an rDNA array that is all actively transcribed by RNA Pol I with a wild type strain containing an rDNA array with half of the repeats actively transcribed and the other half silent. From these experiments we show reduced GpC accessibility suggesting the presence of a bound factor within the upstream element of the 35S rRNA gene promoter.

INTRODUCTION

In *Saccharomyces cerevisiae* the ribosomal DNA (rDNA) locus is ~1-2 Mb tandem array comprised of ~150 to 200 copies of a 9.1 kb repeat on chromosome XII (Petes *et al.*, 1978a; Petes, 1979). A single rDNA repeat consists of two transcribed regions, the 35S rRNA gene transcribed by RNA Pol I and the 5S rRNA gene transcribed by RNA Pol III. Two nontranscribed spacer regions NTS1 and NTS2 surround the 5S rRNA gene (Figure 2a) (Skryabin *et al.*, 1984). RNA Pol I binds the 35S rRNA gene promoter that is formed from an upstream element that stretches from -155 to -60 and a core promoter that stretches from -38 to +5 (Musters *et al.*, 1989; Kulkens *et al.*, 1991; Choe *et al.*, 1992) (Figure 2c). The transcriptional machinery of RNA Pol I consists of Upstream Activation Factor (UAF) (Keys *et al.*, 1996; Siddiqi *et al.*, 2001b) with histones H3 and H4 (Keener *et al.*, 1997), Core Factor (CF) (Keys *et al.*, 1994; Lalo *et al.*, 1996; Lin *et al.*, 1996), Rrn3 (Yamamoto *et al.*, 1996; Milkereit and Tschochner, 1998), and the TATA-binding protein (TBP) (Steffan *et al.*, 1996)(Figure 2c). RNA Pol I utilizes these proteins to form an initiating complex at the 35S rRNA gene promoter.

The promoter contains poised transcriptional machinery that will allow the cell to respond rapidly to a demand for increased ribosome synthesis (Figure 2c). UAF binds to the upstream element to begin assembly of the RNA Pol I initiating complex. Next, core factor and a RNA Pol I-Rrn3 complex are recruited to the promoter. This recruitment requires TBP and is dependent upon UAF bound to the upstream element (Aprikian *et al.*, 2001). Once the pre-initiation complex is formed transcription begins and enters

elongation, TBP, CF, and UAF remain bound to the promoter while RNA Pol I transcribes the 35S rRNA gene (Aprikian *et al.*, 2001; Bier *et al.*, 2004). Rrn3 dissociates from RNA Pol I during elongation, returning to initiate transcription at the 35S rRNA gene promoter by binding RNA Pol I and bridging the interaction between RNA Pol I and CF (Yamamoto *et al.*, 1996; Milkereit and Tschochner, 1998; Peyroche *et al.*, 2000).

Within the rDNA, individual rDNA repeats can exist in two states, either actively transcribed by RNA Pol I with an open chromatin structure characterized by a lack of nucleosomes, or as inactive repeats not transcribed by RNA Pol I with a closed chromatin structure that contains strictly positioned nucleosomes (Dammann *et al.*, 1993, 1995). Active and inactive rDNA repeats have been distinguished using psoralen crosslinking to show that at any one time in an actively dividing cell, approximately half of the 35S rRNA genes are actively transcribed and the other half are silent. Evidence suggests that the transcription state of a 35S rRNA gene is not maintained or inherited from one cell cycle to the next (Dammann *et al.*, 1995; Lucchini and Sogo, 1995). Footprinting assays showed the localization of five well-positioned nucleosomes within NTS2 (Vogelauer *et al.*, 1998) (Figure 2b). Our lab and others have used micrococcal nuclease (MNase) mapping experiments to compare the positions of nucleosomes in the NTS2 region of the rDNA repeats in wild type cells and in cells with a silencing defect caused by deletion of *SIR2* (Fritze *et al.*, 1997; Cioci *et al.*, 2002; Li *et al.*, 2006a). In these studies, changes in positions of nucleosomes in NTS2 were observed in cells lacking the Sir2 protein, a histone deacetylase that is required for silencing of RNA Pol

II transcription (Bryk *et al.*, 1997; Smith and Boeke, 1997) and homologous recombination at the rDNA (Gottlieb and Esposito, 1989). These previous MNase mapping studies within repeated sequences like the rDNA provide the average position of each nucleosome in all of the rDNA repeats. In my research, I sought to determine if a new technique can be utilized to evaluate the accessibility in a single rDNA repeat.

To study the chromatin structure at the single molecule level in the rDNA, a 41 kDa DNA methyltransferase *M.CviPI* is expressed within yeast cells from an estrogen-inducible promoter (Jessen *et al.*, 2004; Hoose and Kladde, 2006). *M.CviPI* originates from a virus that infects Chlorella-like algae, and recognizes GpC sites, methylating the cytosine residue at the C5 position (Xu *et al.*, 1998) (Figure 3c). In chromatin, the methylation of cytosine residues by *M.CviPI* occurs at GpC sites that are located in accessible regions of the genome. Cytosine residues within inaccessible regions of chromatin remain unmethylated (Xu *et al.*, 1998). After purifying genomic DNA, a chemical treatment called bisulfite deamination converts unmethylated cytosines to uracil (Shapiro and Weisgras, 1970; Shapiro *et al.*, 1973; Shapiro *et al.*, 1974) while methylated cytosines remain unchanged, acting as a mark of open, accessible chromatin. After treatment, regions of interest are amplified, cloned into a plasmid and transformed into bacterial cells. Single clones are sequenced and analyzed to distinguish regions of accessible chromatin by GpC sites that retain cytosine from regions of inaccessible chromatin where GpC sites are converted to GT (Jessen *et al.*, 2004; Hoose and Kladde, 2006). Previously published results have shown the value of this single molecule technique in studying changes in positioned nucleosomes from the *PHO5* locus in yeast

cells (Pardo *et al.*, 2009). Using this technique, I analyzed the chromatin structure of individual rDNA repeats to differentiate between open chromatin structures with high levels of GpC methylation from closed chromatin structure with low levels of GpC methylation.

In this study, we used *in vivo* expression of *M.CviPI* to examine changes in chromatin structure at the rDNA NTS2 region due to a silencing defect caused by deletion of *SIR2* and due to RNA Pol I transcription. In wild type cells (MBY1944), we detected high levels of accessibility in three regions: between nucleosome 2 and 3 at the three conserved ARS regions, at a highly methylated region (blue box), and at +4 and +10 within the 35S rRNA gene (Figure 2b). We observed changes in positioning of nucleosome 5, 4, and 3 in NTS2 in *sir2Δ* cells and in cells with a shortened rDNA array (25 rDNA repeats), leading to increased accessibility and GpC methylation. In the 25 copy rDNA array where all 35S rRNA genes are actively transcribed by RNA Pol I, we detected reduced GpC site accessibility suggesting the presence of a bound factor within the upstream element of the 35S rRNA gene promoter. Taken together, our results shed light on how the chromatin structure of NTS2 within the rDNA changes in response to loss of Sir2 and to RNA Pol I transcription.

METHODS AND MATERIALS

Yeast media and strains

Standard yeast media used in these experiments are described elsewhere (Rose, 1990). YPAD consists of YPD media with 40 mg/l of adenine sulfate. All *Saccharomyces cerevisiae* strains are listed in Table 3. Both S288c and W303 yeast strain backgrounds were utilized in this study. Plasmid pSH1052 (aka pSH006a), containing the GpC methyltransferase *M.CviPI* (Xu *et al.*, 1998; Jessen *et al.*, 2004), as well as strains MBY1944 (SHY1860) and its parent strain MBY2076 (Y880/CCY113a) are described elsewhere (Hoose and Kladde, 2006). MBY2076 was used to create MBY2203 by using PCR-mediated gene disruption with a *sir2Δ::KANMX* fragment forming MBY2188, followed by integration of pSH1052 to form MBY2203.

The yeast strain MBY2078 (NOY408-1b *MATa ade2-1 ura3-1 his3-11 trp1-1 leu2-3,112 can1-100*) is described elsewhere (Nogi *et al.*, 1991b). PCR-mediated gene disruption was used to introduce *lys2Δ::*KANMX4 into strain MBY2078 to create MBY2276, and integration of pSH1052 into MBY2276 generated MBY2291. The yeast strain NOY1071 (*MATa ade2-1 ura3-1 trp1-1 leu2-3, 112 his3-11 can1-100 fob1Δ::HIS3* ~25 rDNA copies) is described elsewhere (Cioci *et al.*, 2003). MBY2279 was formed by PCR-mediated gene disruption with *lys2Δ::*KANMX4, followed by integration of pSH1052 into MBY2279 to form MBY2293. PCR-mediated gene disruption was used to introduce the *sir2Δ::HIS3* allele into MBY2291 to create MBY2321.

Table 3. *S. cerevisiae* strains used in this *M.CviPI* study.

Name	Background	Genotype
MBY1944	S288c	*MATa leu2Δ0 lys2Δ0 ura3Δ0 pho3Δ::Rho::pSH006a (ho/L-lexO.pGAL1M.CviPI.ADH1term - LYS2-pADH1-LexA$_{DBD}$.ER$_{HBD}$.VP16$_{AD}$.ADH1$_{term}$-hoR)*
MBY2076	S288c	*MATa leu2Δ0 lys2Δ0 ura3Δ0 pho3Δ::R*
MBY2078	W303	*MATa ade2-1 ura3-1 his3-11 trp1-1 leu2-3,112 can1-100*
MBY2188	S288c	*MATa leu2Δ0 lys2Δ0 ura3Δ0 pho3Δ::R sir2Δ::KANMX4 #1*
MBY2203	S288c	*MATa leu2Δ0 lys2Δ0 ura3Δ0 pho3Δ::R sir2Δ::KANMX4 HO::pSH1052 (ho/L-lexO.pGAL1-M.CviPI-ADH1term – LYS2-pADH1-LexA$_{DBD}$-ER$_{HBD}$-VP16$_{AD}$-ADH1$_{term}$-hoR*
MBY2276	W303	*MATa ade2-1 ura3-1 his3-11 trp1-1 leu2-3,112 can1-100 lys2Δ::KANMX4* #2
MBY2279	W303	*MATa ade2-1 ura3-1 trp1-1 leu2-3, 112 his3-11 can1-100 fob1Δ::HIS3 lys2Δ::KANMX4* ~25 rDNA copies
MBY2291	W303	*MATa ade2-1 ura3-1 his3-11 trp1-1 leu2-3,112 can1-100 lys2 Δ::KANMX4* #2 *ho*::pSH1052 (*ho*/L-*lexO*. *GAL1M.CviPI.ADH1term-LYS2-pADH1-LexA$_{DBD}$.ER$_{HBD}$.VP16$_{AD}$.ADH1$_{term}$-hoR)*
MBY2293	W303	*MATa ade2-1 ura3-1 trp1-1 leu2-3, 112 his3-11 can1-100 fob1Δ::HIS3 lys2Δ::KANMX4 ho*::pSH1052 (*ho*/L-*lexO*. *pGAL1*M.CviPI.*ADH1term-LYS2-pADH1-LexA$_{DBD}$.ER$_{HBD}$.VP16$_{AD}$.ADH1$_{term}$-hoR)* ~25 rDNA copies
MBY2321	W303	*MATa ade2-1 ura3-1 his3-11 trp1-1 leu2-3,112 can1-100 lys2Δ::KANMX4* #2 *ho*::pSH1052 (*ho*/L-*lexO*. *pGAL1*M.CviPI.*ADH1term-LYS2-pADH1-LexA$_{DBD}$.ER$_{HBD}$.VP16$_{AD}$.ADH1$_{term}$-hoR) sir2Δ::HIS3*

Isolation of yeast genomic DNA and bisulfite deamination

Isolation of yeast genomic DNA and bisulfite deamination was carried out as described elsewhere (Jessen *et al.*, 2004; Hoose and Kladde, 2006). For all PCR amplifications using deaminated DNA as a template reaction conditions were optimized, including the annealing temperature against untreated genomic DNA, and carried out as discussed elsewhere (Warnecke *et al.*, 2002; Shen *et al.*, 2007). Oligonucleotides

OM773 and OM774 (Table 4) were used to amplify a ~1,300 bp fragment from NTS2. PCR reactions were extracted with phenol:chloroform:isoamyl alcohol (25:24:1), precipitated in ethanol, resuspended in water, digested with *BamHI* and *PstI* and then ligated into similarly cut pBluescript plasmid (pBST). Ligations were transformed into chemically competent XL-1 Blue cells, and grown for a maximum of 30 min at 37°C in Luria broth to prevent duplication of individual bacterial clones. Transformations were plated onto Luria broth/agarose plates containing 50 mg/L ampicillin, with a final concentration of 0.25 mM IPTG and 0.05 mg/ml X-Gal (Zymo Research) for blue/white screening. Individual colonies were screened for single ligation events by PCR with oligonucleotide primers OM781 and OM782 (Table 4) in order to verify that each plasmid contained one NTS2 fragment. Cultures from verified clones were stored at -70°C in 96-well deep dishes in 25% glycerol.

Sequencing and data analysis

Initial results were obtained using dideoxy sequencing with the SequiTherm Cycle Sequencing Kit using dideoxy CTP (Epicenter Biotechnologies, Madison, WI). A 20 μL colony PCR reaction was used to amplify each NTS2 fragment separately. Reactions were treated with ExoSAP-IT (USB Corporation, Cleveland, OH) to remove unincorporated primers and nucleotides from PCR reactions leaving only the amplified DNA. Sequencing primers (OM448, OM781, OM782, OM807, OM808, OM810, and OM816) were labeled with γ-^{32}P-ATP using T4 Polynucleotide Kinase as described by the SequiTherm Cycle Sequencing Kit protocol (Table 4). In order for the sequencing reaction to proceed, the concentration of dGTP was increased to a final concentration of

4uM. Reactions were carried out as described by the SequiTherm Cycle Sequencing Kit with dideoxy CTP (Epicenter Biotechnologies, Madison, WI) and were separated by gel electrophoresis as described elsewhere (Hoose and Kladde, 2006). Additional data were collected by sending 96-well plates of individual *E. coli* clones in 8% glycerol for sequencing by the University of Florida Interdisciplinary Center for Biotechnology Research (ICBR). Sequences obtained from ICBR were assessed for quality of sequencing and for percent deamination of all lone cytosines, maintaining no less than 97% complete deamination for each clone (Figure 3c).

Data were generated from the NTS2 region of the rDNA from both S288c and W303 *S. cerevisiae* strain backgrounds (Table 5). The number of clones produced indicates the number of NTS2 fragments cloned into the pBST vector, sequenced, and analyzed. Following sequencing, the DNA sequences were aligned against a wild type strand for ease of identifying GpC sites. Alignments were made using the ClustalW program from European Bioinformatics Institute (EBI) (Larkin *et al.*, 2007). Microsoft Excel was used to calculate percent methylation at each possible GpC methylation site. If a site is methylated, the result was recorded as '1' while an unmethylated site where the cytosine is converted to uracil is recorded as '0'.

Table 4. Oligonucleotides used in *M.CviPI* experiments.

Primer	Sequence
OM448	GAGGTGTTATGGGTGGAGGA
OM773	AAAGACGGATCCAATTATAGTTGATTGGATGGGAAATGG
OM774	AAATGACTGCAGCAAATTCATTTCCAAACTCTTTTCAAACTTA
OM781	GTAATACGACTCACTATAGGG
OM782	TCGAGGTCGACGGTATCG
OM807	TTGTGATGTGGAGAATAAGG
OM808	GAAGAGATTTAGTATGTGGG
OM816	GATTGTTTATGTTTTGTGTGATG
OM810	GAATGGTGGTAGTGTAATG

Table 5. Description of data sets generated by *M.CviPI* induction**.**

Strain	Background	Induction Time	Number of clones*
MBY1944	WT S288c	30 min	7
		60 min	27
		90 min	19
		120 min	144
MBY2203	*sir2Δ* S288c	120 min	111
MBY2293	25 copy array W303	120 min	119
MBY2291	WT W303	60 min	114
		90 min	NA
		120 min	101
MBY2321	*sir2Δ* W303	60 min	100
		90 min	NA
		120 min	NA

*NA refers to not analyzed.

RESULTS AND DISCUSSION

Two approaches were taken to analyze the chromatin structure of individual rDNA repeats. Strains were analyzed for changes in nucleosome positioning due to a silencing defect by the loss of Sir2 and due to RNA Pol I transcription. Results were evaluated in two ways, through plotting the percent methylation of each GpC site within NTS2 to visualize the level of methylation between sites and to visualize nucleosome positioning, and, secondly, by using a correlation analysis. This correlation analysis is a statistical method that was used to assess how closely related two sites within a single repeat are to each other with respect to GpC methylation (Figure 10). In these analyses, green refers to anti-correlation, or when one site is methylated the other is not; black indicates no correlation, meaning there is a lack of sufficient variation to provide a statistically significant result; while red indicates a positive correlation, meaning either both sites are methylated or both sites are not methylated. Thus, red represents either open or closed chromatin and can be used to visually represent a change in methyltransferase accessibility, i.e. a change in nucleosome positioning or a change in association of a DNA binding factor. The correlation analyses were performed by Hugh Patterton of the Department of Microbial, Biochemical, and Food Biotechnology, at the University of the Free State, South Africa.

In vitro methylation experiment for site specificity of *M.CviPI*

An initial control experiment was performed to determine if *M.CviPI* had any preference for one methylation site over another. If this were the case, subsequent data

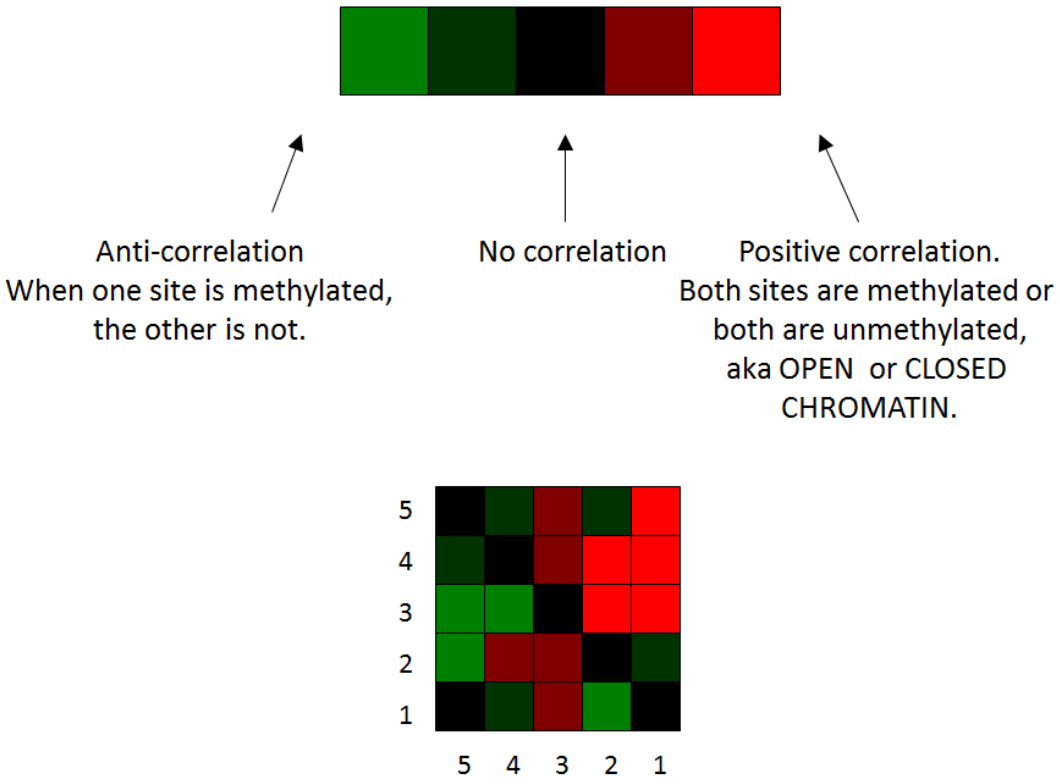


Figure 10. Description of graphical representation of the correlation analyses. Above is an example of the scale of colors produced in these analyses. Green represents anti-correlation, or when one GpC site is methylated the other GpC site is not. Black represents no correlation resulting from a lack of data in this region. In black regions there is not enough data to produce a statistical result. Red represents positive correlation, or when both sites are in agreement. Both are either methylated (and accessible in open chromatin regions) or both are unmethylated (and inaccessible in closed chromatin regions). Thus, red can represent changes in in chromatin structure when comparing two different data sets, for example wild type and sir2Δ. Below is a mock representation of the correlation analyses showing an analysis of five GpC sites. On the X and Y axis the sites are aligned for comparison. A black diagonal line forms in the correlation analysis from comparing a site to itself.

would be skewed toward these sites. Purified naked genomic DNA from the S288c yeast strain MBY1944 was treated *in vitro* with 0.5U of purified *M.CviPI.* DNA was processed and cloned as described in the methods. Thirty-two clones were sequenced to determine if *M.CviPI* favored certain GpC sites in the NTS2 region of the rDNA repeat. Figure 11 shows an almost even level of percent methylation across the NTS2 region, with values ranging between ~15-30%. In this and subsequent graphs, the percent of methylation for each possible GpC site to the total number of clones was calculated. These graphs convert single molecule data into average accessibility data like that obtained with MNase experiments in order to give a simple representation of the data and to easily visualize positioned nucleosomes to scale within NTS2. Methylation peaks at GpC site -725 but subsequent experiments showed that site -725 is not overly favored by *M.CviPI* perhaps because it is occluded from *M.CviPI* by nucleosome 3. In addition, the graph does not show protection patterns that would be expected if the DNA was contaminated with positioned nucleosomes meaning that the DNA was truly 'naked' and thus there was no interference from DNA bound factors. To further analyze these results, a correlation analysis was performed (Figure 12). There seems to be two regions that differ in the intensity of the correlation within NTS2; however, this is likely due to the two separate sequencing reactions that were needed to span the NTS2 region in each clone. With that in mind, the correlation analysis on the naked DNA samples shows a relatively even accessibility between all sites, indicating that *M.CviPI* has little to no site specificity in the NTS2 DNA sequences.

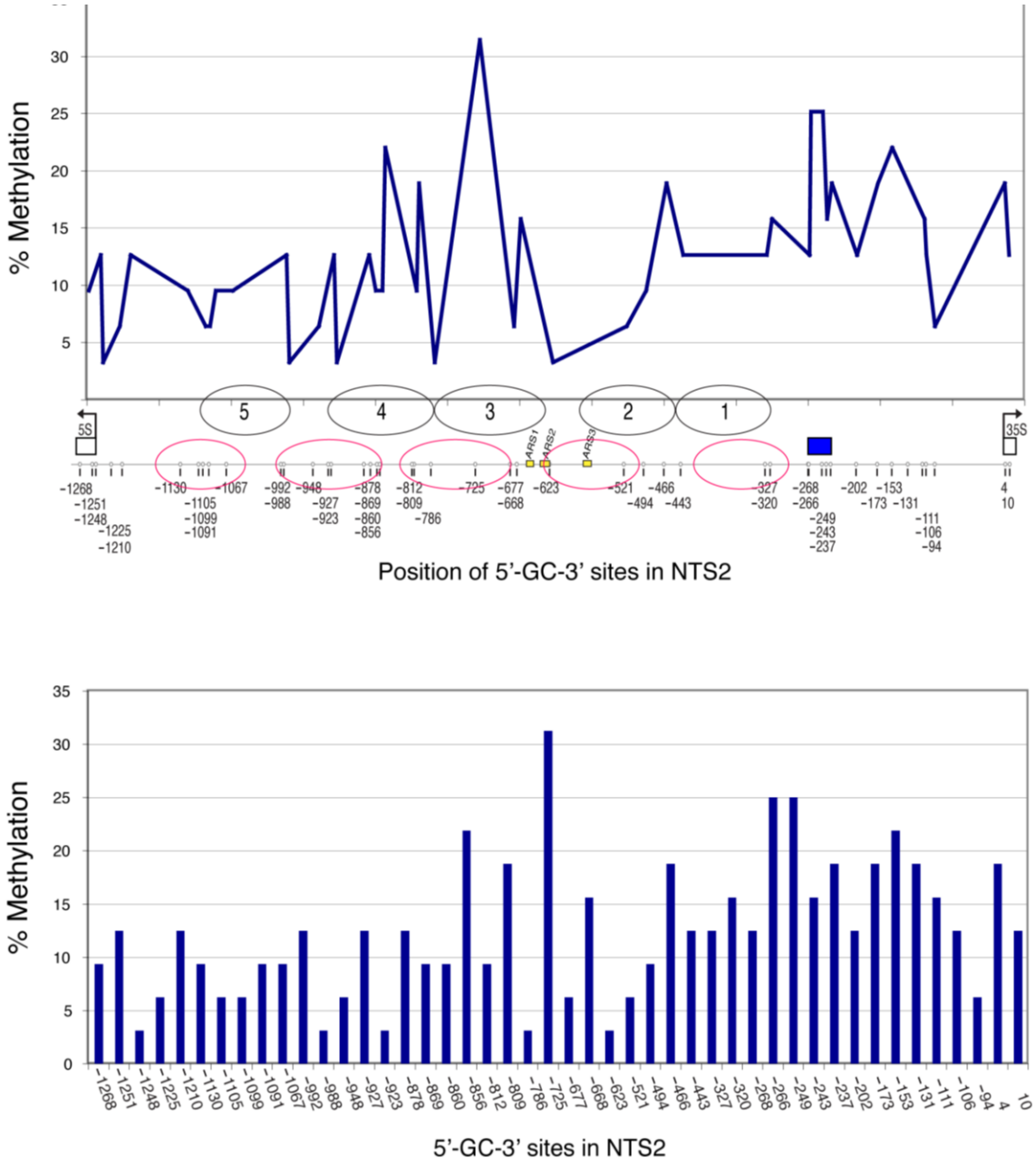


Figure 11. *In vitro* methylation of naked DNA from S288c background MBY1944 treated with 0.5U of *M.CviPI*. In vitro methylation was performed to determine site specificity of *M.CviPI* in NTS2 on naked DNA. Total percent methylation out of the total number of sites available was calculated. **(Top)** This graph shows nucleosome positioning and GpC sites to scale. The five positioned nucleosomes shown as numbered circles outlined in black are based upon MNase mapping (Vogelauer *et al.*, 1998; Li *et al.*, 2006a). Circles outlined in red show nucleosome positioning from *M.CviPI* accessibility. Each GpC is numbered on the x-axis relative to the transcriptional start site (+1) of the 35S rRNA gene. The 5S and 35S rRNA genes are indicated with arrows and white boxes to the left and right, respectively. Yellow boxes represent conserved ARS sequences 1, 2, and 3. A blue box represents a highly methylated region. **(Bottom)** This bar graph shows the same data as above, except it is not to scale with the GpC sites.

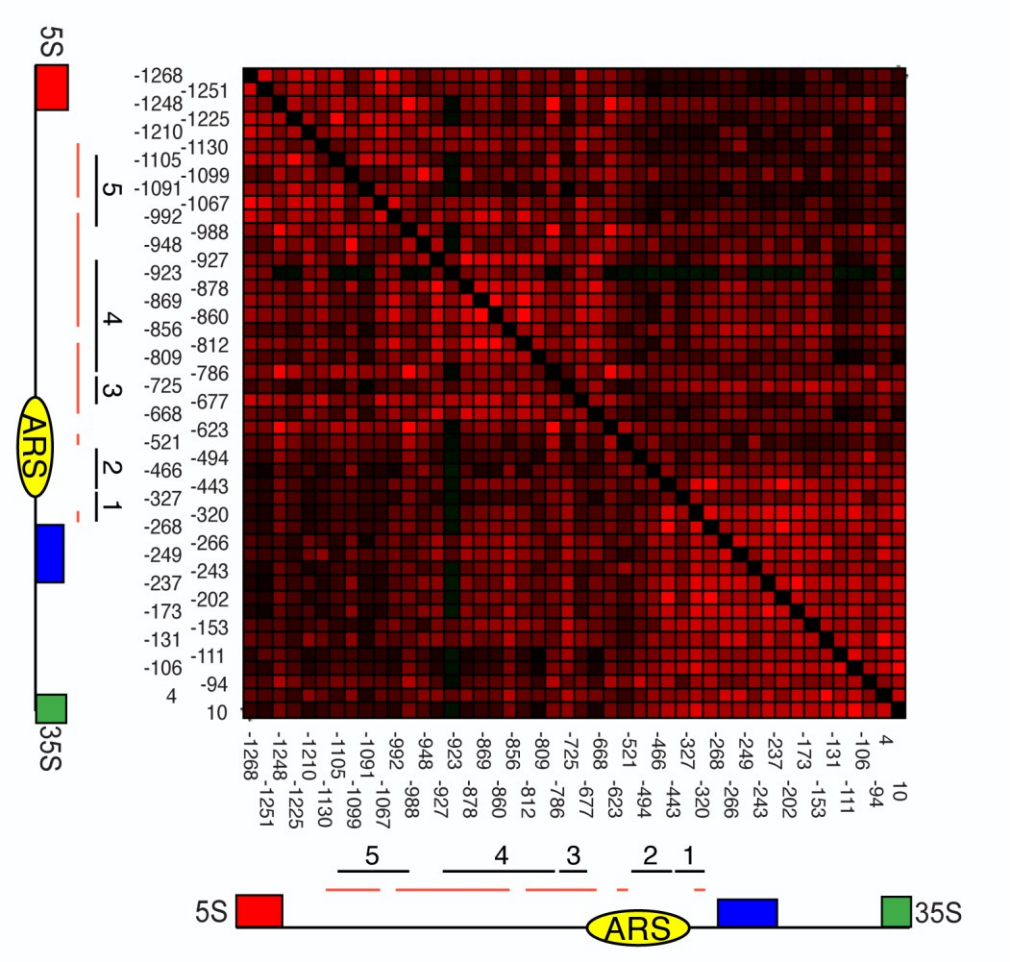


Figure 12. Correlation analysis of *in vitro* methylation of naked DNA from S288c background MBY1944 treated with 0.5U of *M.CviPI.* On the X and Y axis lie the 5'-GpC-3' sites within the NTS2 region. A red box and green box represent the 5S and 35S rRNA genes, respectively. A yellow circle encompasses the conserved ARS sequences. A blue box represents the highly methylated region. Note that the image is not to scale. In these analyses, green refers to anti-correlation, or when one site is methylated the other is not; black indicates no correlation, meaning there is a lack of sufficient variation to provide a statistically significant result; while red indicates a positive correlation, meaning either both sites are methylated or both sites are not methylated. Black and red lines above the representation of the NTS2 region represent the GpC sites where nucleosomes 1-5 lie within NTS2. Black represents the nucleosome positions from previous data (Li *et al.*, 2006a) while red lines represent nucleosome positioning from the data presented in this chapter.

Time course of *M.CviPI* induction in S288c yeast strain MBY1944

Initial experiments focused on optimizing the induction time of *M.CviPI.* Inducing for too short or too long of a time may result in under or over methylation of the DNA, respectively. A time course of 30, 60, 90, and 120 minutes was performed using a wild type S288c yeast strain MBY1944 (Figure 13). From this induction trial, 120 minutes was deemed optimal for DNA methylation. The 30 and 60 minute induction of *M.CviPI* DNA methylation was too short to obtain complete data on all positioned nucleosomes. This is evident when looking at the correlation analysis in Figure 14. In these experiments only a short region of NTS2 was analyzed. For the 30, 60, and 90 minute correlation analyses there is insufficient variation to provide statistically significant correlation, i.e. the copious amount of black regions. Only in the clones from the 120-minute induction do we see clear positioning of all five nucleosomes (Figure 13) and a finer scale of correlation between sites (Figure 14).

When choosing a technique to study nucleosome positioning, the size of the probe can affect the accuracy of how well the location of DNA-bound factors associated with chromatin is mapped. Smaller probes are able to get closer to DNA-bound factors than probes that are larger; therefore, smaller probes have the potential to provide a higher resolution map. *M.CviPI* is a 41 kDa protein while MNase is 17 kDa and psoralen is the smallest with a three-ringed furocoumarin structure with the molecular formula $C_{11}H_6O_3$. Of the three, *M.CviPI* is the largest and admittedly not the best probe in terms of size; however, with each probe there are pros and cons. By expressing *M.CviPI in vivo*, it can be utilized for single molecule accessibility to study chromatin

structure within a living cell, while both MNase and psoralen are *ex vivo* probes that provide average accessibility data. From our experiments with *M.CviPI*, we see differences in positioned nucleosomes in NTS2 similar to previously published results obtained using MNase (Li *et al.*, 2006a). In Figure 13, nucleosomes outlined in black represent the previously published MNase mapping data, while nucleosomes outlined in red represent approximate positions determined from the data generated with *M.CviPI*. The differences in the positions are likely due to the size of the probe. These data do not disprove previous positioning data (Vogelauer *et al.*, 1998; Li *et al.*, 2006a). These differences are simply a caveat of using *M.CviPI*. The data presented here are also relevant and conclusions can be made about changes in nucleosome positioning and open/inactive repeats by analysis of single molecules.

Recent experiments using FISH have theorized that during an entire cell cycle a single repeat in a single cell is active for five minutes before turning off and that during that one cycle every repeat will eventually switch on (Tan and van Oudenaarden, 2010). Our experiments were based on the idea that an actively transcribed repeat remains in an active state throughout the cell cycle. Using our technique, a 5-minute induction time would be too short to produce enough methylation in NTS2. The induction times of 60 and 120 minutes of *M.CviPI* are necessary for this technique based on the correlation analyses produced during the induction time course series (Figures 13, 14).

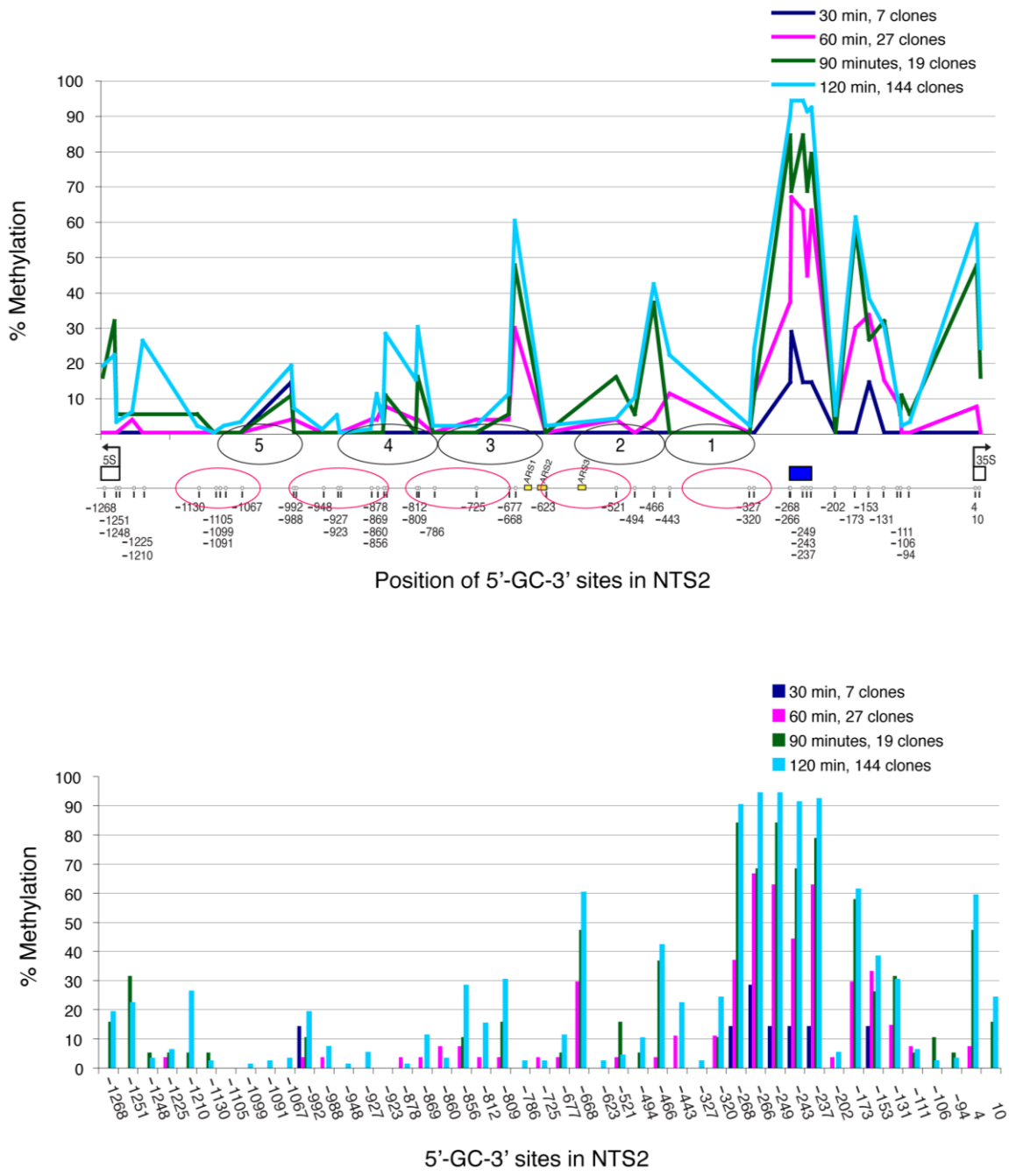


Figure 13. Percent methylation in time course induction of *M.CviPI* in S288c yeast strain MBY1944. An induction of *M.CviPI* with estrogen for 30, 60, 90, and 120 minutes is shown with the total number of clones for each listed at the top right of the figure. Refer to Figure 11 for further description of this figure.

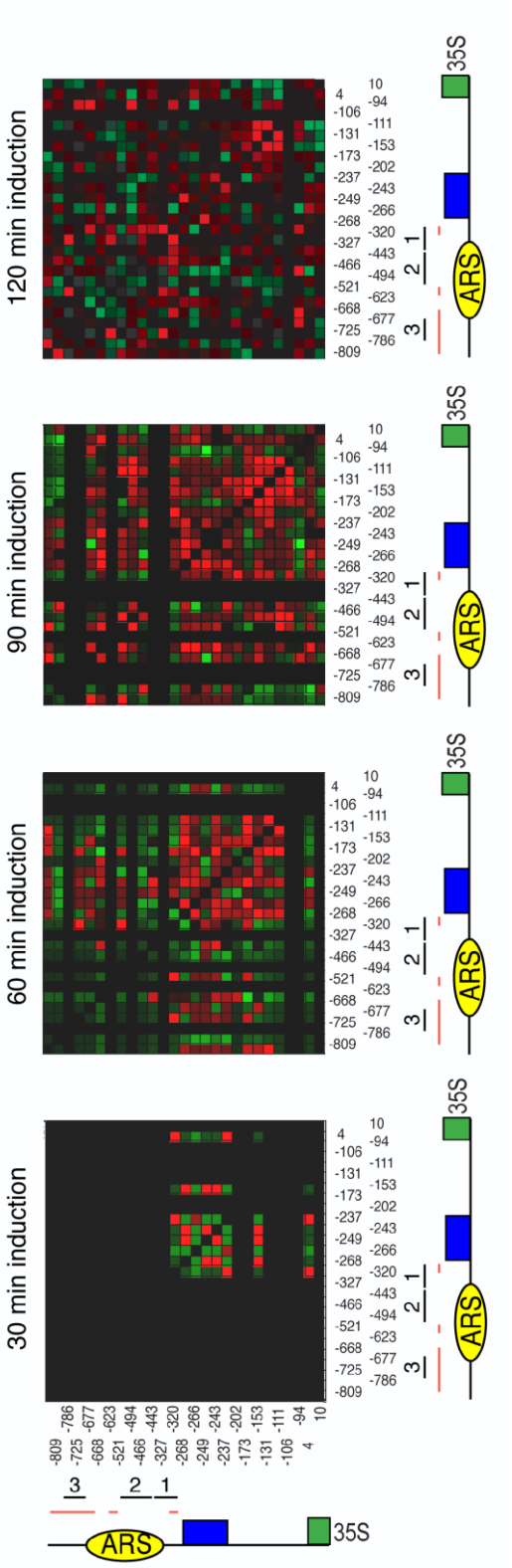


Figure 14. Correlation analysis of time course of induction in S288c yeast strain MBY1944. Refer to Figure 12 for a description of this figure

Methylation in wild type and *sir2Δ* S288c strains after 120 min induction

Our lab previously showed changes in nucleosome positioning within NTS2 by performing MNase mapping experiments in cells with a silencing defect due to deletion of *SIR2* (Li *et al.*, 2006a). In cells lacking Sir2, increased MNase accessibility was seen between each nucleosome at linker regions downstream of nucleosome 1, between nucleosome 2 and 3 representing the ARS region, and upstream of nucleosome 5 (Cioci *et al.*, 2002; Li *et al.*, 2006a). We sought to utilize expression of *M.CviPI in vivo* for single molecule analysis of nucleosome positioning in wild type and *sir2Δ* cells. Using this single molecule approach we hoped to discover new characteristics of the chromatin structure within NTS2 that could not be seen with MNase, in addition to visualizing results similar to what has been previously observed using MNase. The classic way of presenting single-molecule methylation data is with a dot diagram (Figure 15). A dot diagram was created with a random selection of 50 NTS2 clones from S288c wild type yeast cells (MBY1944) after a 120-minute induction of *M.CviPI*. In dot diagrams, black dots represent a GpC site that was methylated by *M.CviPI* while white dots represent a GpC site that was not methylated by *M.CviPI*. We observed three regions with high levels of accessibility in wild type cells: at -668 and -667 between nucleosome 2 and 3 near ARS1 and ARS2 conserved sequences, downstream of nucleosome 1 at a region designated by a blue box, and within +4 and +10 of the 35S rRNA gene. These data from wild type cells are compared to *sir2Δ* data in Figure 16, revealing few differences between the two data sets.

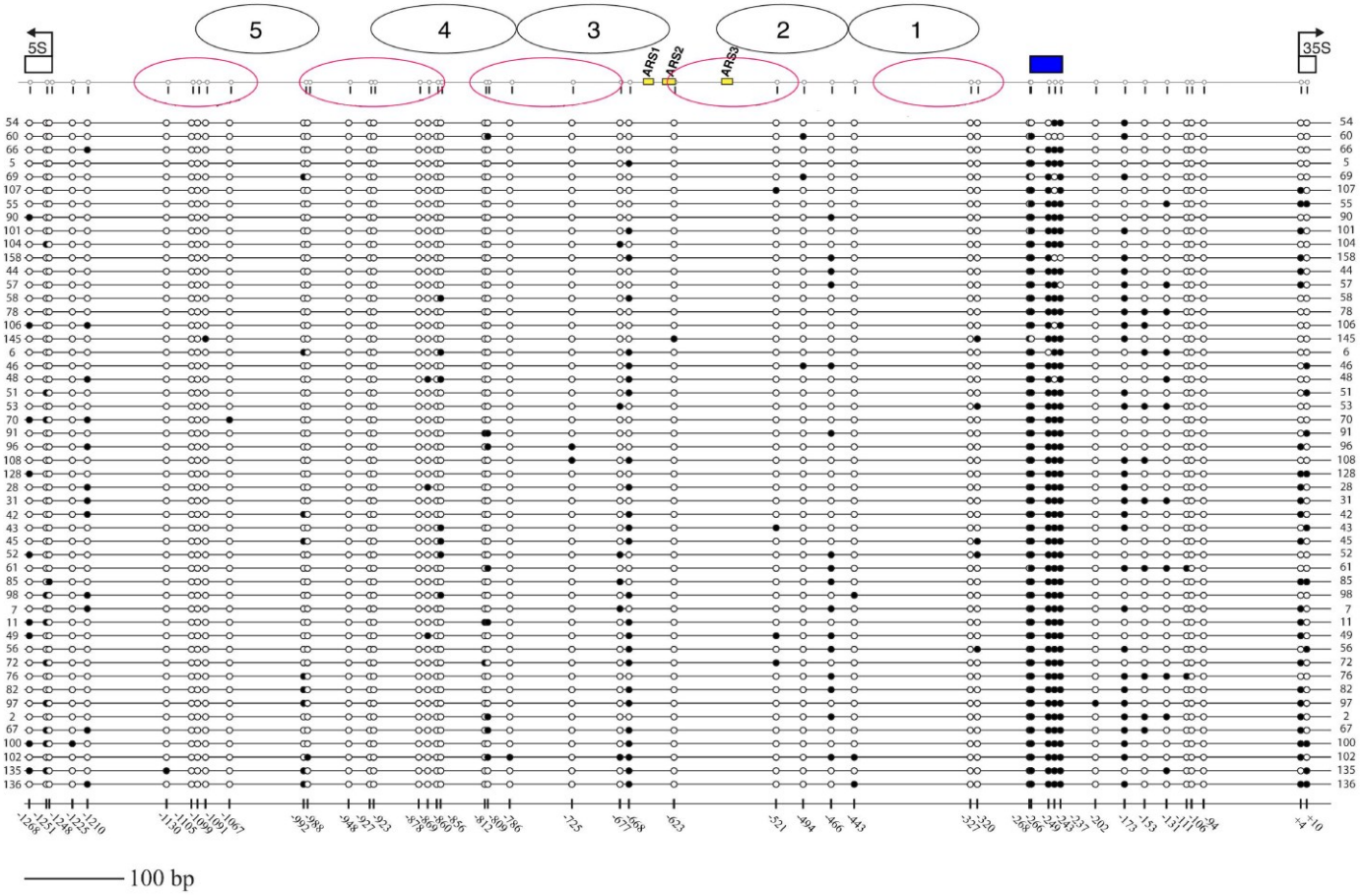


Figure 15. Dot diagram of wild type MBY1944 clones in S288c yeast background after a 120 minute induction. Refer to Figure 11 for further description of this figure. Black dots represent a GpC site that was methylated by *M.CviPI* while white dots represent a GpC site that was not methylated by *M.CviPI*. Values listed to the left and right of each clone represent a specific clone number.

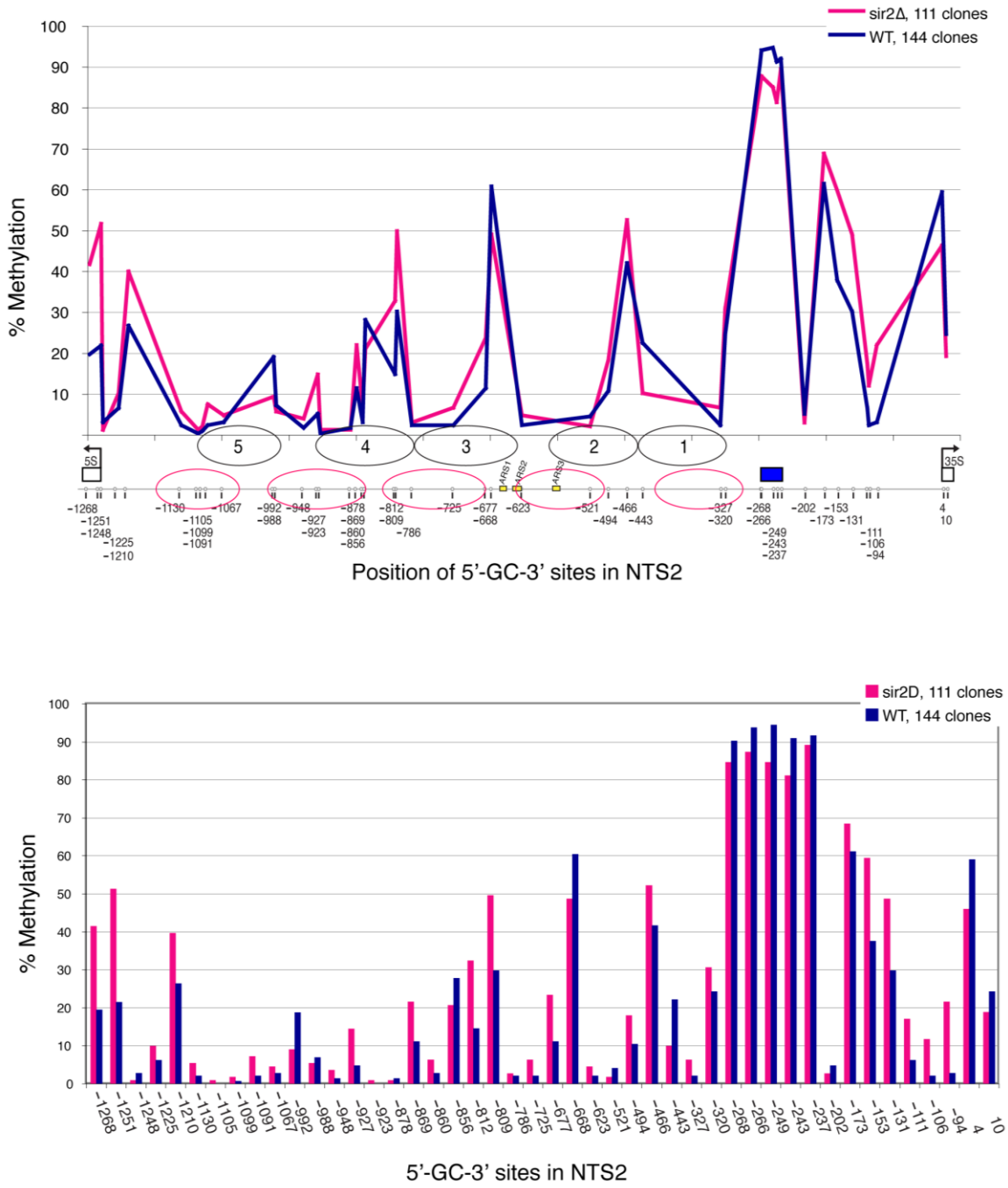


Figure 16. Percent methylation in wild type (MBY1944) and *sir2Δ* (MBY2203) in S288c background after 120 minute induction. The total number of clones are listed at the top right of the figure. Refer to Figure 11 for a description of this figure.

In the correlation analysis for these data, there is an overall increase in positive/red correlation suggesting a change in chromatin structure has occurred in *sir2Δ* cells (Figure 17). Remember that this correlation analysis is a statistical method that assesses how closely related two sites within a single repeat are to each other with respect to GpC methylation. Positive/red correlation can indicate that either both sites are methylated or both sites are not methylated, meaning either open (accessible) or closed (inaccessible) sites. This can be interpreted as a change in the positioning of a DNA bound factor. In comparing the region of nucleosome 5 and 4 between wild type and *sir2Δ* data, we see a region of negative/black correlation in wild type cells is changed to positive/red correlation in *sir2Δ* cells. This suggests that in *sir2Δ* cells there is an increase in accessibility to *M.CviPI* and thus a change in the positions of these nucleosomes. Differences are also seen within the highly methylated region, though no factor is known to bind in this region. Also, changes can be seen within the 35S rRNA gene promoter region at GpC site -106, an area within the upstream element (Figure 2c). We expected to see differences in accessibility within the ARS region. In cells lacking Sir2 there is an increase in DNA replication and ARS firing at the rDNA locus (Pasero *et al.*, 2002), but from this analysis we cannot detect any obvious changes in methyltransferase accessibility in *sir2Δ* cells. It could be that increased origin activity in *sir2Δ* cells is not associated with changes in *M.CviPI* accessibility.

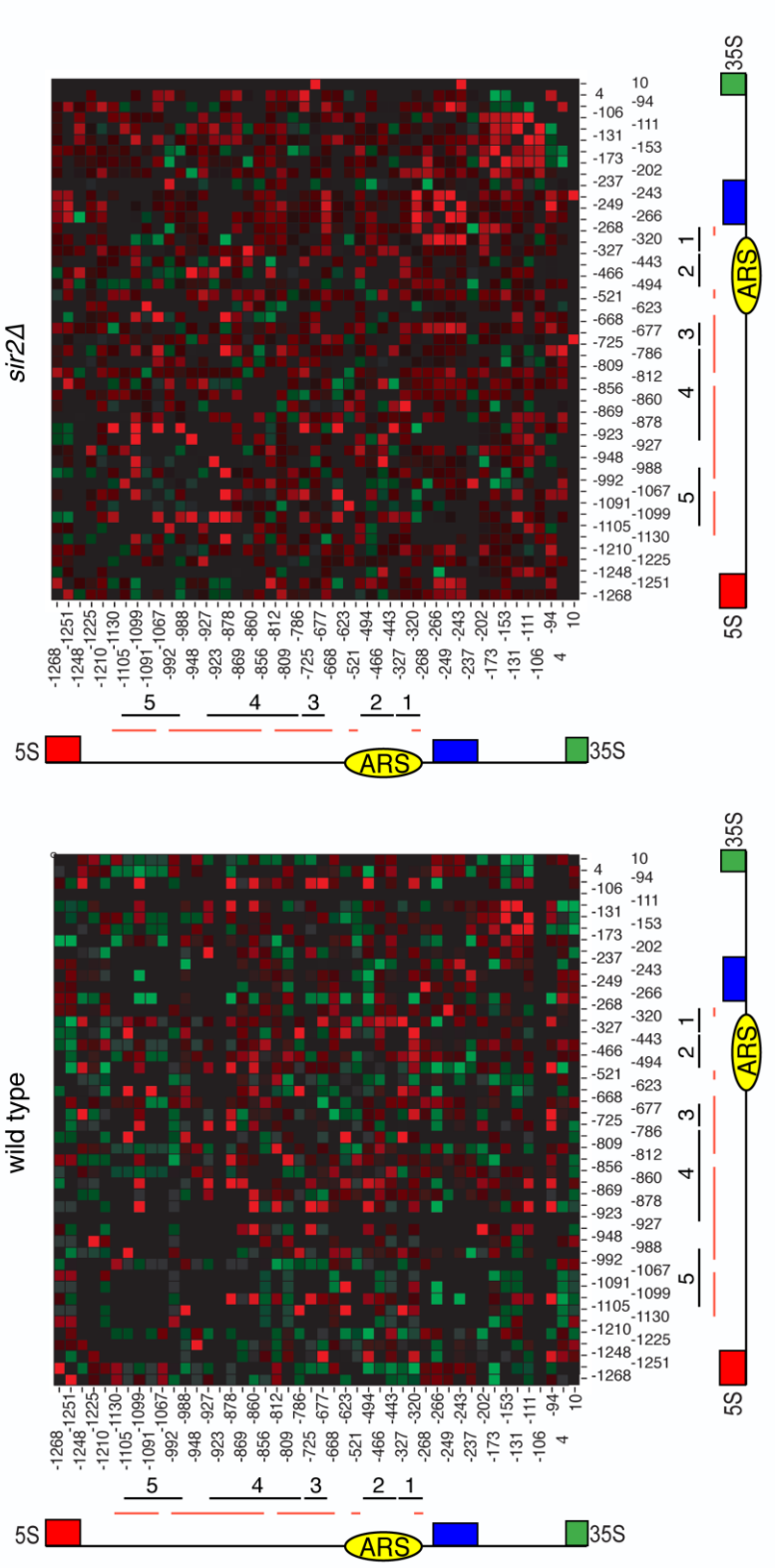


Figure 17. Correlation analysis of wild type (MBY1944) and *sir2Δ* (MBY2203) in S288c background after 120 minute induction.
Refer to Figure 12 for a description of this figure.

Methylation of wild type and *sir2Δ* in W303 strains after 60 min induction

M.CviPI was induced in wild type (MBY2291) and *sir2Δ* (MBY2321) W303 strains for 60 minutes. As discussed above, a 60-minute induction time was not optimal; however, with a decreased induction time we aimed to obtain a clearer correlation analysis. When graphed, our ability to visualize positioned nucleosomes is decreased, but we can see changes between wild type and *sir2Δ* cells (Figure 18). Downstream of nucleosome 1 there are clear changes in percent methylation between wild type and *sir2Δ* cells within the highly methylated region. We see the opposite of what is expected, with decreased accessibility (decreased methylation) in *sir2Δ* cells compared to wild type.

Changes in nucleosome positioning can be seen clearly in the correlation analysis (Figure 19). The correlation graph from the wild type clones clearly shows negative/black correlation at regions within nucleosome 5, 4, and 3. In *sir2Δ* clones, positions of nucleosome 5, 4, and 3 have shifted. There is an increase in negative/black correlation around the region of nucleosome 5 and 4, and nucleosome 3 has shifted downstream. These nucleosomes occupy a larger area in *sir2Δ* cells, suggesting less well defined nucleosome positioning. In addition, analysis of the wild type clones revealed a large amount of positive/red correlation at the highly methylated region and the 35S rRNA gene promoter region, which is decreased in *sir2Δ* cells to an overall negative/green correlation. Differences in these regions are visible in the percent methylation graph as well. Alterations in the accessibility of the NTS2 region in

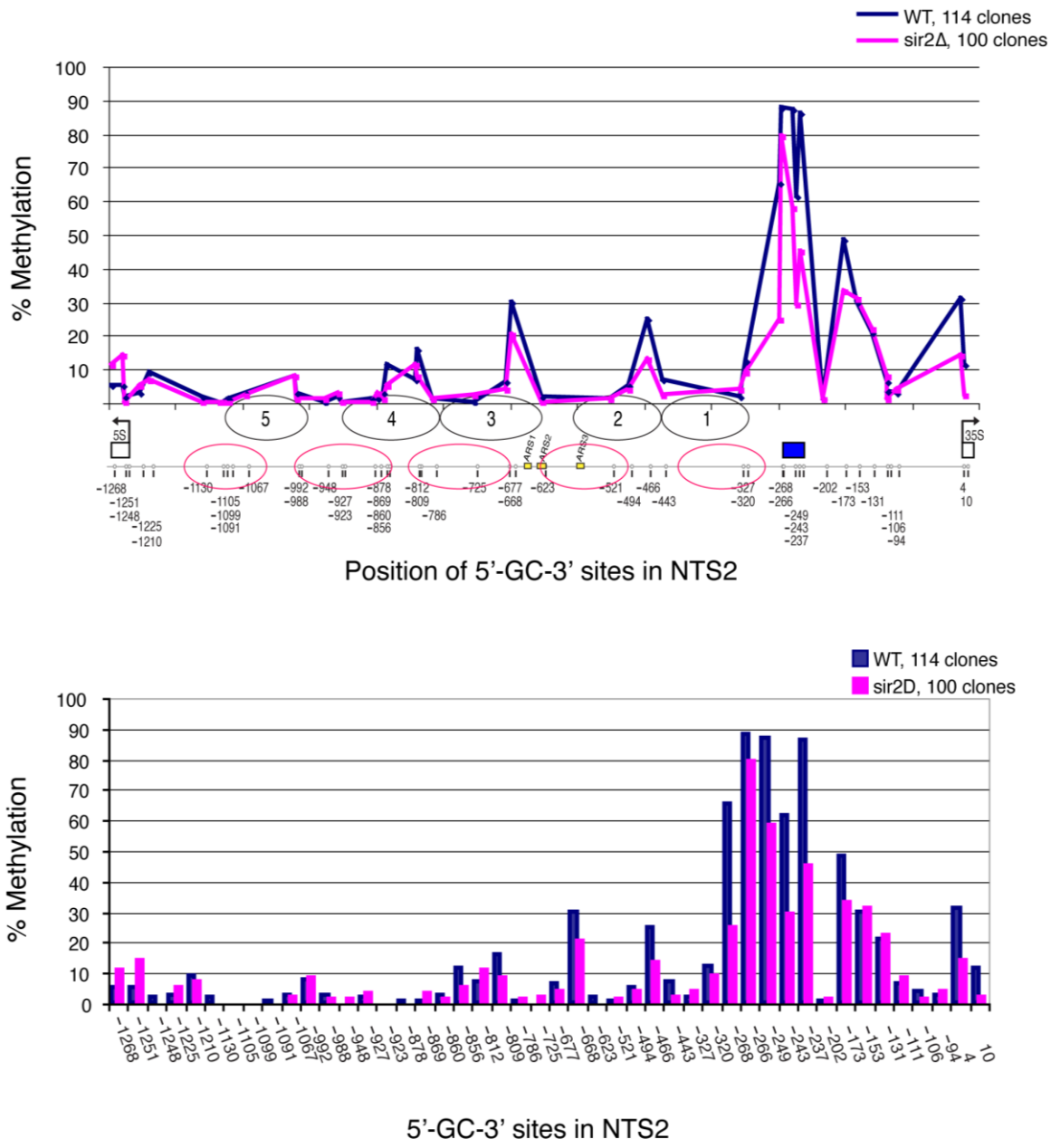


Figure 18. Percent methylation in wild type (MBY2291) and *sir2Δ* (MBY2321) in W303 background after 60 minute induction. The total number of clones are listed at the top right of the figure. Refer to Figure 11 for a description of this figure.

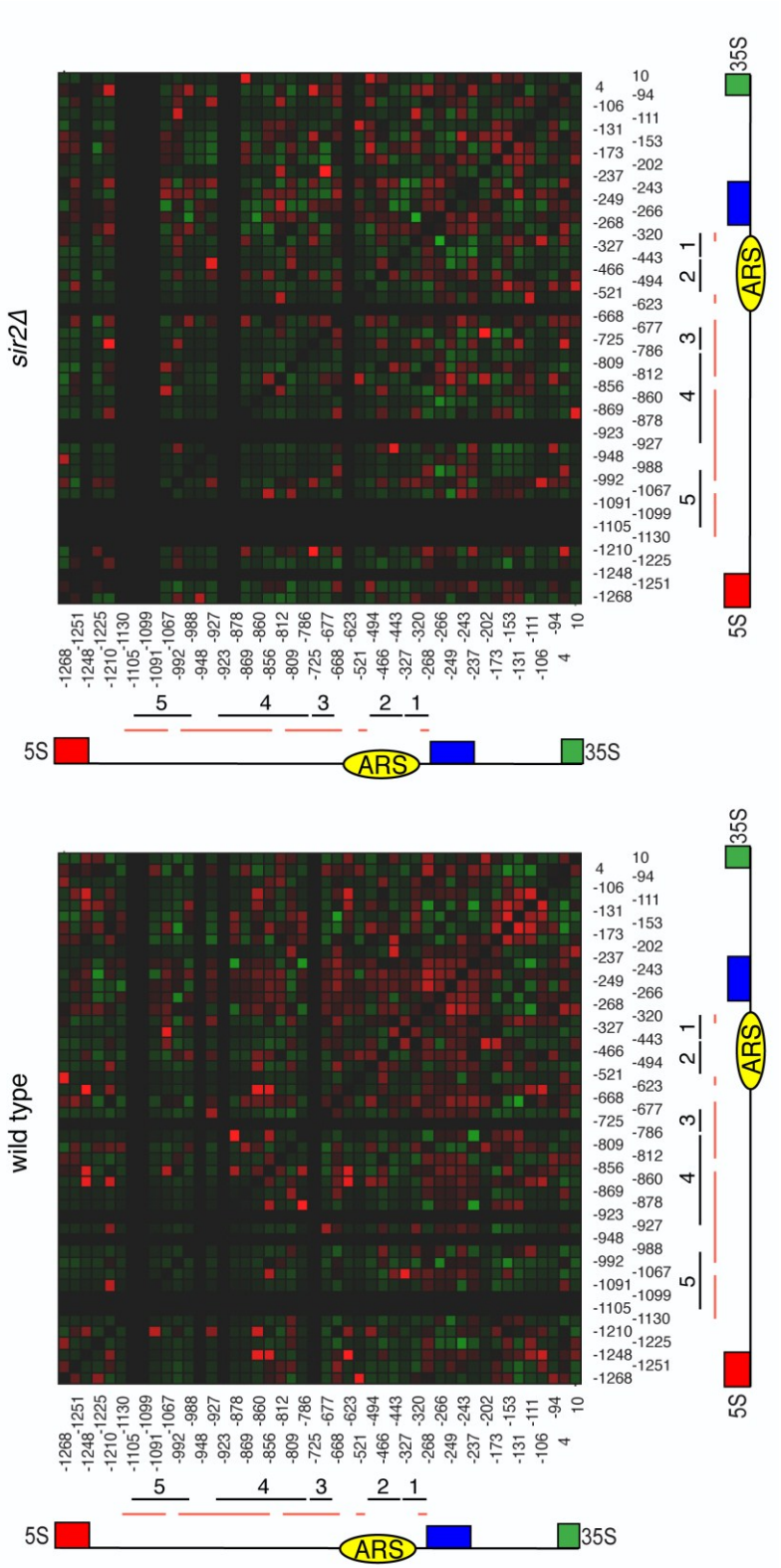


Figure 19. Correlation analysis of wild type (MBY2291) and *sir2Δ* (MBY2321) in W303 background after 60 minute induction. Refer to Figure 12 for a description of this figure.

wild type and *sir2Δ* cells to *M.CviPI* indicate a change in chromatin structure at the highly methylated region and the 35S rRNA gene promoter region.

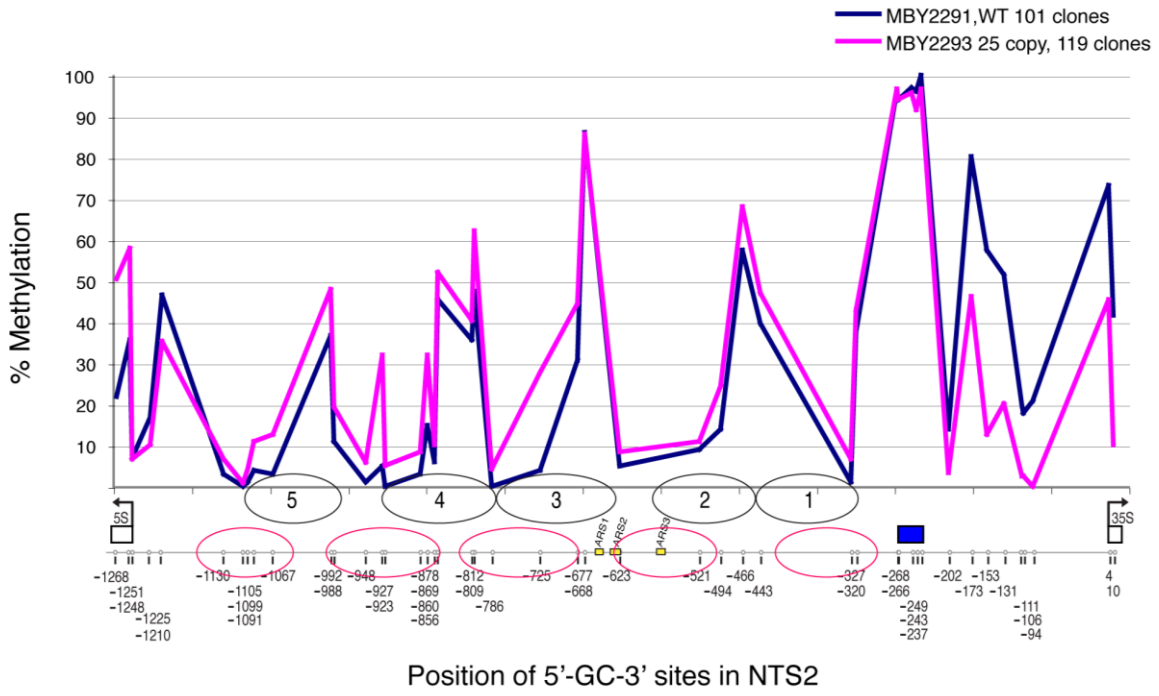
By combining a graph of percent methylation and correlation analysis we detect changes in nucleosome positioning in *sir2Δ* cells similar to previously published data obtained using MNase (Cioci *et al.*, 2002; Li *et al.*, 2006a). Overall, nucleosomes 5, 4, and 3 exhibit changes in positioning in *sir2Δ* cells, characterized by an increase in black/negative correlation at 60 minute induction in Figure 19 and an increase in positive correlation at 120 minutes in Figure 17. This difference in data is due to the different induction times of *M.CviPI* used in these experiments. With longer induction times, there will be less black/negative correlation due to a lack of variation in the methylation data and thus by looking at both 60 and 120 minute inductions we can see changes in nucleosome positioning in *sir2Δ* cells. Unlike with MNase data, we did not see changes between nucleosome 2 and 3, and downstream of nucleosome 1 in *sir2Δ* cells.

Methylation of 150-200 rDNA repeats and 25 rDNA repeats in W303 strains after 120 min induction

Our second approach to studying the chromatin structure of individual rDNA repeats was to analyze for different populations of rDNA repeats based on RNA Pol I transcription. In wild type cells that contain ~150-200 rDNA repeats, approximately half of the 35S rRNA genes are transcribed by RNA Pol I (Dammann *et al.*, 1993). We expected actively transcribed repeats to be more accessible to *M.CviPI* because RNA Pol I-transcribed genes have been shown to be free of nucleosomes or non-nucleosomal (Dammann *et al.*, 1993, 1995). To facilitate identification of actively transcribed

repeats, we analyzed *M.CviPI* accessibility in a strain that carries a 25 rDNA copy array (MBY2293) with all twenty five 35S rRNA genes transcribed by RNA Pol I and compared the results to those from wild type cells (MBY2291). Our experiments were based on the hypothesis that an actively transcribed 35S rRNA gene remains active throughout a single cell cycle.

Within the 35S rRNA gene promoter region, we detected reduced methylation, representing the binding of RNA Pol I transcriptional machinery (Figure 20). The region surrounding -94 GpC site represents binding of RNA Pol I transcriptional machinery at the upstream element in the 35S rRNA gene promoter region (Figure 2c). The correlation analysis has a predominance of black or no correlation, meaning a lack of variation among the rDNA repeats analyzed. Again, this suggests the binding of a DNA binding factor that blocks *M.CviPI* accessibility (Figure 21). UAF is bound at the upstream element of the 35S rRNA promoter at all times, regardless of whether or not the gene is transcribed by RNA Pol I (Claypool *et al.*, 2004). However, Hmo1, a HMG-box protein, associates exclusively with actively transcribed 35S rRNA genes within the promoter region (Kasahara *et al.*, 2007; Merz *et al.*, 2008). Our data suggests that we are able to detect the association of Hmo1 with actively transcribed 35S rRNA genes. Hmo1 is an interesting factor because of its similarity to human UBF (Schnapp *et al.*, 1994). In mammalian cells, UBF binds to the upstream control region of the 45S rRNA promoter through its HMG boxes forcing ~140 bp into a 360° loop (Bazett-Jones *et al.*, 1994; Stefanovsky *et al.*, 2001). Hmo1 contains two HMG box domains, however it



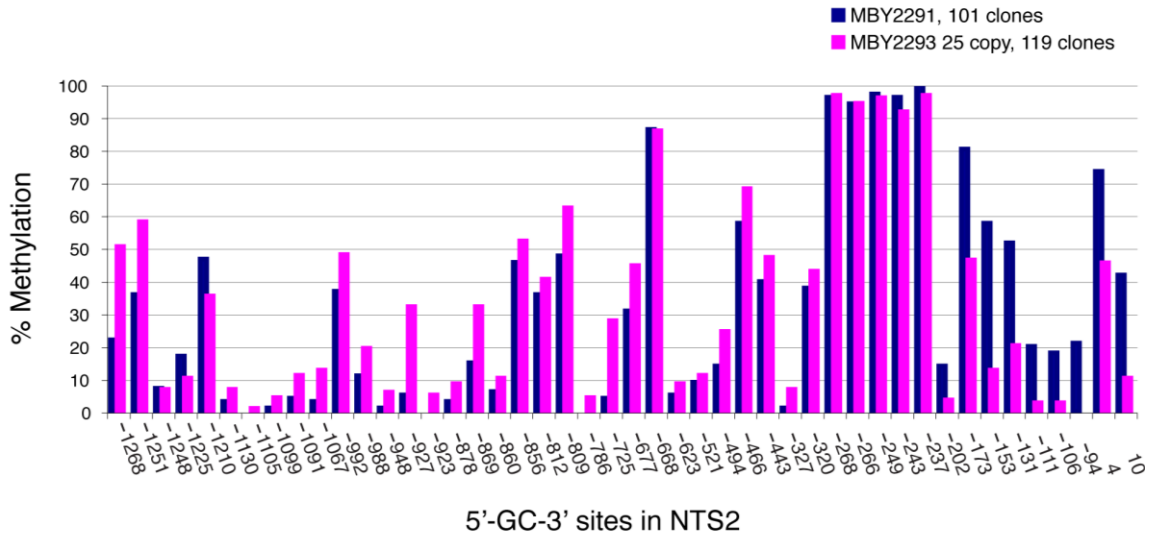


Figure 20. Percent methylation in wild type (MBY2291) and 25 copy array (MBY2293) in W303 background after 120 minute induction. The total number of clones are listed at the top right of the figure. Refer to Figure 11 for a description of this figure.

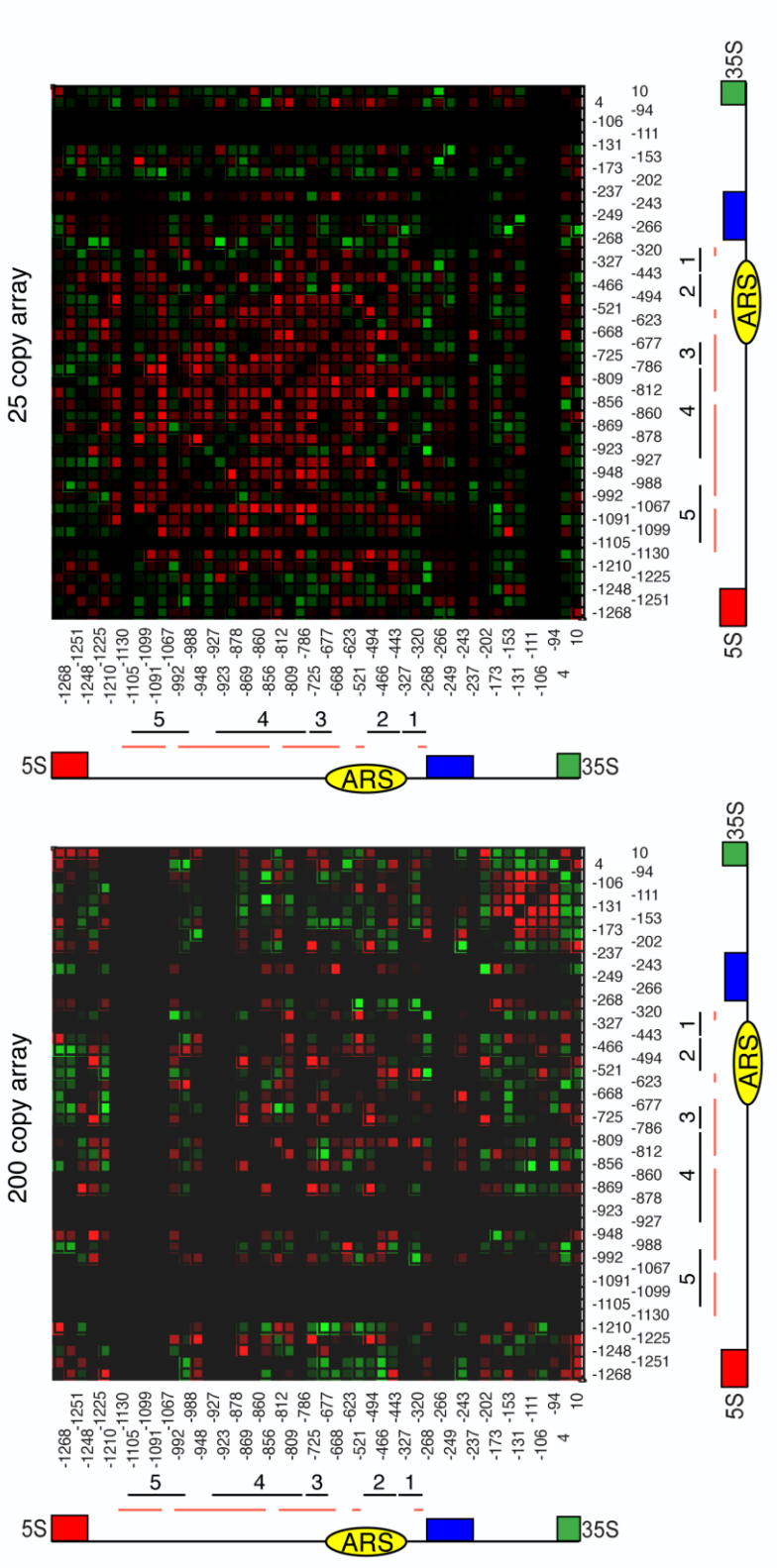


Figure 21. Correlation analysis of wild type (MBY2291) and 25 copy array (MBY2293) in W303 background after 120 minute induction. Refer to Figure 12 for a description of this figure.

uses its C-terminus to bind and bend the DNA but no evidence of looping has been observed (Xiao *et al.*, 2010).

In the data from the wild type clones, much of the black correlation can be attributed to strict nucleosome positioning, in particular nucleosomes 5, 4, and 3. Much of the negative correlation observed in the data from wild type clones is replaced with color in the data from the 25 copy cells, suggesting that in a 25 copy rDNA array nucleosomes are less well positioned or not present, which would account for greater variation in the level of methylation that was observed. Actively transcribed regions of the rDNA may contain either unstable nucleosomes or none at all due to being constantly displaced by RNA Pol I transcriptional machinery (Dammann *et al.*, 1993, 1995; Jones *et al.*, 2007; Merz *et al.*, 2008). The correlation analysis in Figure 21 represents the analysis of only a portion of the total data produced in these experiments: 18 clones from the wild type 200 copy array cells and 37 clones from the 25 copy array cells. The lack of correlation in the highly methylated region (blue box, Figure 21) is unique to this strain background. In other analyses using the S288c background, we did not detect a lack of correlation at this highly methylated region. This difference could be due to the sample size analyzed or to the binding of an unknown factor.

We used a single molecule technique to analyze changes in nucleosome positioning within NTS2 of the rDNA due to a silencing defect caused by deletion of Sir2 and due to active transcription by RNA Pol I. In wild type and *sir2Δ* experiments, we observed changes in the positioning of nucleosomes 5, 4, and 3, but did not detect differences in accessibility either between nucleosome 2 and 3 or downstream of

nucleosome 1 that has been observed previously in MNase experiments (Figure 17, 19) (Cioci *et al.*, 2002; Li *et al.*, 2006a). In order to further analyze specific nucleosomes, analysis of *M.CviPI* accessibility should be carried out on smaller regions of NTS2 rather than the whole fragment at once. These windows of data could focus on smaller areas of specific nucleosomes, on the ARS region, the highly methylated region, and on the 35S rRNA promoter region to visualize changes in accessibility. With wild type and 25 copy experiments, our data suggests the presence of a bound factor within the 35S rRNA promoter that associates exclusively with actively transcribed 35S rRNA genes within the upstream element (Figure 20, 21). Based on published articles it is likely that this factor is Hmo1 discussed in Chapter I (Merz *et al.*, 2008). Future work should focus on identifying factors bound within active repeats, whether it is Hmo1 or any other factors present at the promoter region of actively transcribed 35S rRNA genes. These correlation analyses are an unproven technique and although conclusions have been made about the 25 copy clones containing a bound factor unique to actively transcribed repeats, these conclusions should be verified with a more traditional technique. ChIP could easily be used to verify the presence of a bound factor that is enriched within the RNA Pol I promoter region in actively transcribed genes.

Taken together, this work has analyzed changes in chromatin structure using the DNA methyltransferase *M.CviPI* as a probe. We measured *M.CviPI* accessibility comparing wild type cells to those with a silencing defect due to deletion of *SIR2* and to cells with fewer rDNA repeats. We detected changes in nucleosome positioning in NTS2 and changes at the 35S rRNA promoter, likely due to the association of RNA Pol I

factors. Our results show that alterations in silent chromatin structure at the rDNA can be detected using a single molecule technique that takes advantage of *M.CviPI*.

CHAPTER III

CHARACTERIZATION OF CONSERVED AMINO ACIDS IN THE SET DOMAIN OF *S. cerevisiae* Set1[3]

In *S. cerevisiae*, the protein methyltransferase Set1 is part of a larger complex called COMPASS (Complex associated with Set1) that associates with RNA Polymerase II. Set1 contains a conserved SET domain with an active site where substrates are methylated using the methyl group from S-adenosyl methionine (SAM). Mutants of Set1 were made within the SET domain to learn more about the catalytic mechanism of Set1. The crystal structures of human SET domain proteins, as well as sequence alignments and a random mutagenesis of yeast Set1 were used to identify conserved amino acids in the SET domain of Set1. Here we present Set1 mutants characterized with quantitative Western blots to measure the levels of the different forms of K4-methylated H3, and with quantitative Northern blots to evaluate gene silencing at the ribosomal DNA locus. Set1 has also recently been found to perform a role in chromosome segregation through methylation of Dam1. Set1 mutants have been characterized for their ability to suppress an *ipl1-2* temperature sensitive phenotype. Set1 mutants R1013H and Y967F have been tested for their ability to form the COMPASS complex *in vivo* using tandem affinity purification (TAP) from yeast cell extracts. TAP purification of Set1 with R1013H and Y967F mutations has been used in

[3] John Mueller contributed to quantitative Westerns measuring the level of mono-, di-, and trimethylated H3K4 relative to either Pgk1 or total histone H3 (Figure 21, Table 8). Victoria Schneider contributed the *ipl1-2* plate assays (Table 9).

in vitro assays for their ability to methylate lysine 4 of histone H3. This analysis will increase understanding of the catalytic mechanism and cellular function of *S. cerevisiae* Set1 and provide insight into Set1-like proteins in higher eukaryotes.

INTRODUCTION

Histone methylation of lysine residues has long been studied for its role in both euchromatin and heterochromatin regions, specifically in RNA Pol II transcription, heterochromatin formation, gene silencing, and chromosome segregation. The majority of histone methylation is carried out by proteins that contain a highly conserved SET domain, its origins from *Drosophila* genes Suvar 3-9, Enhancer of zeste, and Trithorax (Jones and Gelbart, 1993; Tschiersch *et al.*, 1994). SET domain proteins have been found in a wide range of organisms from simple model systems all the way to mouse and human systems. With each step up on the evolutionary ladder, SET domain proteins and their roles within those systems become increasingly complex. In humans, SET domain proteins have been linked to a multitude of diseases such as leukemia, breast cancer and colon cancer (Ziemin-van der Poel *et al.*, 1991; Prasad *et al.*, 1997; Natarajan *et al.*, 2010).

SET domain proteins can modify lysine residues of histone H3 up to three times using S-adenosyl methionine (SAM) as the methyl donor. Mono-, di-, and trimethylation of lysine 4 of histone H3 by Set1 has been linked to high levels of gene expression as well as silencing of genes (Nislow *et al.*, 1997; Briggs *et al.*, 2001; Roguev *et al.*, 2001; Bryk *et al.*, 2002; Krogan *et al.*, 2002; Santos-Rosa *et al.*, 2002). Our lab

has focused on studying methylation of H3K4 at the ribosomal DNA locus in *S. cerevisiae*, one of three heterochromatin regions. Set1 is part of a multi subunit complex called COMPASS (Complex associated with Set1) that associates with RNA Pol II through the Paf1 complex (Briggs *et al.*, 2001; Miller *et al.*, 2001; Roguev *et al.*, 2001). The subunits of COMPASS contain counterparts in human H3K4 methyltransferase complexes, like the MLL complexes and Set1A and Set1B complexes. Although yeast Set1 is predominantly studied for its role in actively transcribed genes, it was first discovered for its role in silencing transcription at telomeres, mating type loci, and the rDNA (Nislow *et al.*, 1997; Briggs *et al.*, 2001; Bryk *et al.*, 2002; Krogan *et al.*, 2002). At the rDNA, RNA Pol II transcription is silenced despite high levels of RNA Pol I and RNA Pol III transcription. With deletion of *SET1*, there is a loss of silencing of RNA Pol II transcription, suggesting a model where Set1 is required for silencing of transcription at the rDNA (Bryk *et al.*, 2002). In *sir2Δ* cells there is loss of silencing of RNA Pol II transcription and an increase in di- and trimethylated H3K4 within NTS2 (Li *et al.*, 2006a). It is not yet clear what form(s) of H3K4 methylation (mono-, di-, or trimethylated) is required for gene silencing in *S. cerevisiae* or how the low levels of H3K4 methylation found at the rDNA alter silent chromatin function. In addition to methylated H3K4, Set1 also methylates Dam1, a protein involved in chromosome segregation during the cell cycle. Methylation of Dam1 at K233 leads to proper chromosome segregation by blocking phosphorylation of a neighboring serine residue by Ipl1 aurora kinase. Phosphorylation halts chromosome segregation when microtubules

improperly attach to the kinetochore and is part of the anaphase checkpoint during the cell cycle (Zhang *et al.*, 2005; Westermann *et al.*, 2007).

In humans, there are six known groups of histone methyltransferases that methylate H3K4: MLL complexes 1-5, Set1A and Set1B complexes, Set7/9, Ash1, Smyd3, and Meisetz (Malik and Bhaumik, 2010) (Appendix A). Each of these complexes are involved in specific and complex roles within human cells. Set7/9 forms the majority of our understanding of the catalytic mechanism of SET domain proteins due to its small size and because it functions without any additional subunits (Wilson *et al.*, 2002; Kwon *et al.*, 2003; Xiao *et al.*, 2003). Recent crystal and *in vitro* analysis of MLL1 have also provided insight into the catalytic mechanism of SET domain proteins, especially yeast Set1 as it is more similar to MLL1 than Set7/9 (Patel *et al.*, 2009; Southall *et al.*, 2009). Computer simulations with the crystal structure of Set7/9 have focused on the importance of tyrosine residues within the catalytic SET domain to maintain an ordered hydrogen bonding with water molecules present inside a substrate specific channel with lysine 4 of histone H3 that is critical for deprotonation of SAM and thus methylation of lysine 4 (Zhang and Bruice, 2007, 2008a). Because of the multiple H3K4 methyltransferases available in humans, study of these complexes is incredibly difficult. This chapter describes a mutational analysis of yeast Set1 that was performed to identify and characterize amino acid residues in the SET domain of Set1 that are required for the methylation of histone H3 at K4 and transcriptional silencing in the rDNA. In our model system *S. cerevisiae*, Set1 is the sole H3K4 methyltransferase

allowing a straightforward characterization *in vivo*, easily providing information about other SET domain proteins in eukaryotes.

Specific amino acids within the catalytic SET domain of *SET1* were chosen from an alignment of SET domain proteins and by studying the crystal structures of Set7/9 and MLL1. Set1 mutants were created by site-directed mutagenesis as well as random mutagenesis, and were characterized using quantitative Westerns for H3K4 methylation *in vivo* and Northerns for transcriptional silencing of a RNA Pol II-transcribed gene at the rDNA. Furthermore, methylation of Dam1 was analyzed using *ipl1-2* temperature sensitive strains. Mutants were also assessed for their effect on COMPASS formation using Tandem affinity purification followed by *in vitro* assays with increasing concentrations of SAM. Overall, we have identified residues in three categories of mutants: mutants with wild type activity, partially active mutants, and null mutants.

MATERIALS AND METHODS

Media

Standard yeast media used in these experiments are described elsewhere (Rose, 1990). YPADT consists of YPD media with 40 mg/l of adenine sulfate and 20 mg/l of l-tryptophan.

Plasmids and mutagenesis

Plasmid MBB251 contains the *SET1* gene flanked by a *Xho1* site and 436 bp upstream and a *SacII* site and 347 bp downstream of the *SET1* ORF. MBB251 was modified by the addition of a *Cla1* site 9 bp downstream of the SET1 stop codon to make MBB484. The *Xho1*-*SacII* fragment of MBB484 was ligated into pRS406 to create MBB491 to create a vector for integration of *SET1* sequences into *ura3-52* (see below). The Bluescript plasmid was modified by cloning a linker containing a *Mun1* restriction site into the unique *EcoR1* site to create plasmid MBB479. For mutagenesis, the *Mun1*-*Cla1* fragment of *SET1* from MBB484 was ligated into MBB479 to make MBB487. Set1 mutants were generated by site-directed mutagenesis with Phusion polymerase (New England Biolabs, Ipswich, MA) and specific primers (Table 6). Mutagenized sequences were verified by DNA sequence analysis (Gentech lab and Interdisciplinary Center for Biotechnology Research, University of Florida). For each mutant, the *Mun1*-*Cla1* fragment was ligated into the *Mun1*-*Cla1* interval of MBB491 to create full length mutant *SET1* in pRS406.

Table 6. Oligonucleotides used in Set1 experiments.

Primer	Description	Sequence
OM336	TAP primer	TGGGATATAATCGATGAGTTGGAGCAAATTGCC CTACAGCAATCCATGGAAAAGAGAAG
OM337	TAP primer	ACACTTTCAGTGTGTTTTAATTATTCTTCTTTGA ATGCTGCTTACGACTCACTATAGGG
OM486	N1016A sense	GGTGGTATAGCCCGTTTCATTGCTCATTGTTGTG ATCCAAATTGTACGGC
OM487	N1016A antisense	GCCGTACAATTTGGATCACAACAATGAGCAATG AAACGGGCTATACCACC
OM504	G951A sense	CGTTCAGCAATTCACAACTGGGCTTTATATGCT CTAGACTCTATCGC
OM505	G951A antisense	GCGATAGAGTCTAGAGCATATAAAGCCCAGTTG TGAATTGCTGAACG
OM506	Y967A sense	GCAAAGGAAATGATTATCGAGGCCGTTGGTGA AAGGATCAGGC
OM507	Y967A antisense	GCCTGATCCTTTCACCAACGGCCTCGATAATCA TTTCCTTTGC
OM508	I972A sense	CGAGTACGTTGGTGAAAGGGCCAGGCAACCTG TAGCAG
OM509	I972A antisense	CTGCTACAGGTTGCCTGGCCCTTTCACCAACGT ACTCG
OM510	Y993A sense	GGGATTGGATCCAGTGCCCTTTTTAGGGTTGAT GAAAACACGG
OM511	Y993A antisense	CCGTGTTTTCATCAACCCTAAAAAGGGCACTGG ATCCAATCCC
OM512	H1017L sense	GGTATAGCCCGTTTCATTAATCTTTGTTGTGATC CAAATTGTACGGC
OM513	H1017L antisense	GCCGTACAATTTGGATCACAACAAAGATTAATG AAACGGGCTATACC
OM636	T1024Y sense	CATTAATCATTGTTGTGATCCAAATTGTTATGCA AAGATTATAAAGGTTGGCGGGAG
OM637	T1024Y antisense	CTCCCGCCAACCTTTATAATCTTTGCATAACAAT TTGGATCACAACAATGTAAAATG
OM638	R1013H sense	GCCACCAAGAAAGGTGGTATAGCCCATTTCATT AATCATTGTTGTGATCCA
OM639	R1013H antisense	TGGATCACAACAATGATTAATGAAATGGGCTAT ACCACCTTTCTTGGTGGC
OM708	Cla1 primer	GGTTTCTTGAACTGACGATGATACATCGATTTG TTTGGAGCTTCCTG
OM709	Cla1 primer	CAGGAAGCTCCAAACAAATCGATGTATCATCGT CAGTTCAAGAAACC
OM712	SET1_FWD	CCTGAACTTTGTCAAAGAGAAGAATCCTCG
OM713	SET1_RVS	GAATGCTGTCGGTAGCCTAAAGAAATGCTTC

Table 6. continued.

Primer	Description	Sequence
OM749	Y967F sense	GGAAATGATTATCGAGTTCGTTGGTGAAAGGAT CAGG
OM750	Y967F antisense	CCTGATCCTTTCACCAACGAACTCGATAATCAT TTCC
OM751	Y1054F sense	GCGAAGAGTTGACATATGATTTCAAATTTGAGA GAGAAAAGG
OM752	Y1054F antisense	CCTTTTCTCTCTCAAATTTGAAATCATATGTCAA CTCTTCGC
OM753	T1024A sense	CATTGTTGTGATCCAAATTGTGCGGCAAAGATT ATAAAGG
OM754	T1024A antisense	CCTTTATAATCTTTGCCGCACAATTTGGATCACA ACAATG
OM755	Y1054A sense	GAAGAGTTGACATATGATGCCAAATTTGAGAG AGAAAAG
OM756	Y1054A antisense	CTTTTCTCTCTCAAATTTGGCATCATATGTCAAC TCTTC
OM757	F1056Y sense	GACATATGATTACAAATATGAGAGAGAAAAGG ATGACGAG
OM758	F1056Y antisense	CTCGTCATCCTTTTCTCTCTCATATTTGTAATCA TATGTC
OM851	Mun1 linker	TGATGCCATGAATTCGTAGCAATTGGAATTCTA GGTGCAT
OM852	Mun1 linker	ATGCACCTAGAATTCCAATTGCTACGAATTCAT GGCATCA
OM947	F1056A sense	GAAGAGTTGACATATGATTACAAAGCTGAGAG AGAAAAGGATGACGAG
OM948	F1056A antisense	CTTCTCAACTGTATACTAATGTTTCGACTCTCTC TTTTCCTACTGCTC
OM949	I989A sense	GAGAAAAGATATCTGAAAAATGGGGCTGGATC CAGTTACCTTTTTAGG
OM950	I989A antisense	CCTAAAAAGGTAACTGGATCCAGCCCCATTTTT CAGATATCTTTTCTC
OM951	H1017A sense	GCCCGTTTCATTAATGCTTGTTGTGATCCAAATT GTACG
OM952	H1017A antisense	CGTACAATTTGGATCACAACAAGCATTAATGAA ACGGGC
OM953	H1017R sense	GCCCGTTTCATTAATCGTTGTTGTGATCCAAATT GTACG
OM954	H1017R antisense	CGTACAATTTGGATCACAACAACGATTAATGAA ACGGGC
OM955	Y1052F sense	GCGGCAAGCGAAGAGTTGACATTTGATTACAA ATTTGAGAGAG

Table 6. continued.

Primer	Description	Sequence
OM956	Y1052F antisense	CTCTCTCAAATTTGTAATCAAATGTCAACTCTTC GCTTGCCGC
OM957	Y1052A sense	GCGGCAAGCGAAGAGTTGACAGCTGATTACAA ATTTGAGAG
OM958	Y1052A antisense	CTCTCAAATTTGTAATCAGCTGTCAACTCTTCGC TTGCCGC
OM959	Y1052V sense	GCGGCAAGCGAAGAGTTGACAGTTGATTACAA ATTTGAGAG
OM960	Y1052V antisense	CTCTCAAATTTGTAATCAACTGTCAACTCTTCGC TTGCCGC
OM961	E1048A sense	CTGCGTGATATCGCGGCAAGCGCTGAGTTGACA TATGATTAC
OM962	E1048A antisense	GTAATCATATGTCAACTCAGCGCTTGCCGCGAT ATCACGCAG
OM963	E1049A sense	CTGCGTGATATCGCGGCAAGCGAAGCTTTGACA TATGATTAC
OM964	E1049A antisense	GTAATCATATGTCAAAGCTTCGCTTGCCGCGAT ATCACGCAG
OM965	L1050A sense	GATATCGCGGCAAGCGAAGAGGCTACATATGA TTACAAATTTGAG
OM966	L1050A antisense	CTCAAATTTGTAATCATATGTAGCCTCTTCGCTT GCCGCGATATC
OM967	R1042H sense	AGAATTGTTATCTATGCACTGCATGATATCGCG GCAAGCG
OM968	R1042H antisense	CGCTTGCCGCGATATCATGCAGTGCATAGATAA CAATTCT
OM969	R1042A sense	AGAATTGTTATCTATGCACTGGCTGATATCGCG GCAAGCG
OM970	R1042A antisense	CGCTTGCCGCGATATCAGCCAGTGCATAGATAA CAATTCT
OM971	C1068A sense	GACGAGGAAAGACTTCCTGCTTTATGTGGAGCA CCTAATTGT
OM972	C1068A antisense	ACAATTAGGTGCTCCACATAAAGCAGGAAGTCT TTCCTCGTC
OM973	C1070A sense	GAGGAAAGACTTCCTTGTTTAGCTGGAGCACCT AATTGTAAAGG
OM974	C1070A antisense	CCTTTACAATTAGGTGCTCCAGCTAAACAAGGA AGTCTTTCCTC
OM975	C1075A sense	TGTTTATGTGGAGCACCTAATGCTAAAGGTTTC TTGAACTGACG
OM976	C1075A antisense	CGTCAGTTCAAGAAACCTTTAGCATTAGGTGCT CCACATAAACA

Set1 mutants generated by random mutagenesis were created by PCR from wild type *SET1* gene in MBY1198, described previously (Bryk *et al.*, 2002), using primers OM712 and OM713 (Table 5) to amplify the SET domain through two rounds of PCR (Hirschhorn *et al.*, 1995). The SET domain fragment was digested with *Cla1* and *Mfe1* and inserted in to a similarly cut MBB484 containing full length *SET1* using gap repair into MBY2269. Mutants were screened based on transcription of a RNA Pol II-transcribed gene *Ty1his3AI.* Set1 mutants that resulted in a loss of silencing at the rDNA will become HIS^+. Mutants were analyzed by plating aliquots of 5mL YPADT cultures of each mutant grown at 20°C for four days onto SC-Ura agarose plates for total number of cells compared to cells plated on SC-Ura-His plates. The ratio to a wild type SET1 strain and a set1Δ strain was calculated and used as a basis for judging randomly generated Set1 mutants. Random mutants were selected for a partial loss of silencing phenotype and are listed in Table 7. Only E1049G and C1075Y were analyzed further.

Table 7. Summary of Set1 mutants generated by random mutagenesis of the SET catalytic domain.

Random mutant	Description
RM 2	F995Y
RM 6	A959T, I1003Y
RM 8	E1049G
RM 13	S945L
RM 14	I958T, E1049G
RM 26	I1015Y, T1051A
RM 29	N1016D
RM 44	L1066P
RM 46	E924G
RM 47	G990E
RM 56	K982R, F995L
RM 62	I972T
RM 72	D1053G
RM 74	R1065K
RM 80	Y967C
RM 98	I964F, D1004G
RM 118	Y1039C
RM 122	C1075Y
RM 141	G969C
RM 142	Y993H, F1056L
RM 145	Y967H

Yeast strains

All *Saccharomyces cerevisiae* strains are listed in Table 8. Yeast strains were made by using standard genetic crosses and transformation techniques. Wild type and mutant alleles of *SET1* in the integrating vector were digested with *Stu1* and transformed into MBY2269 and ZK2 *Δset1*. Integration of a single copy of MBB491 with wild-type *SET1* or derivatives containing mutant alleles of *SET1* into the *ura3-52* locus were verified by PCR and Southern blot analysis. All yeast strains containing Set1 mutants were verified to be a single copy integrated into *ura3-52* by PCR by southern blot analysis.

Yeast strain MBY2269 was generated from a cross between MBY1206 (*MATa ura3Δ0 leu2Δ0 his3Δ200 ade2Δ::hisG trp1Δ63 met15Δ0* Ty1*his3AI*-236 Ty1*ade2AI*-515) and MBY2046 (*MATα ade2Δ::hisG his3Δ200 leu2-(Δ0orΔ1) met15Δ0 trp1Δ63 ura3-52 set Δ::TRP1* Ty1*his3AI-236 ADE2::TELVR)* and verified by marker checks and PCR. Strains made using MBY2269 were used in quantitative Western analysis and northern analysis. Strains used in the *ipl1-2* plate assays were made with MBY2450 and have been described elsewhere as ZK2 *Δset1* (Zhang *et al.*, 2005). Set1 mutants were integrated into *ipl1-2* strains and were verified to be a single copy integrated into *ura3-52* by PCR and southern blot analysis. Strains used in tandem affinity purification (TAP) were made with MBY2469, with the originating strain SF10 (BJ5459 *MATa ura3-52 trp1Δ lys2-801 leu2Δ1 pep4::HIS3 prb1Δ1 can1*-) described elsewhere (Li *et al.*, 2002). The addition of sequences encoding the TAP tag to the C-terminus of the yeast *BRE2* gene was performed as described in Puig et al. (1998) Yeast 14, 1139-1146.

Oligonucleotides OM336 and OM337 were used to amplify a fragment containing sequences encoding the TAP tag and *K. lactis TRP1* gene flanked by sequences with homology to the carboxy terminus of the *BRE2* coding region. The fragment was transformed into yeast strain BJ5459 (MBY1212) and Trp+ transformants were selected. After identification of strains with the correct integration of TAP sequences at the C-terminus of the *BRE2* gene, the *SET1* gene was replaced by a *set1Δ::KANMX4* fragment as described (Bryk *et al.*, 2002) to create yeast strain MBY2469. Strains used in tandem affinity purification (TAP) were made with MBY2469, with the originating strain described elsewhere (Li *et al.*, 2002).

Table 8. *S. cerevisiae* strains used in this Set1 study.

Name	Genotype
BJ5459	*MATa ura3-52 trp1Δ lys2-801 leu2Δ1 pep4::HIS3 prb1Δ1 can1*$^{-}$
MBY2269	*MATa ade2Δ::hisG his3Δ200 leu2- (Δ0 or Δ1) met15Δ0 trp1Δ63 ura3-52 set1Δ::TRP1* Ty1*his3AI*-236 *ADE2::TELVR*
MBY2313	MBY2269 *ura3-52:: URA3*-pRS406-*SET1*
MBY2315	MBY2269 *ura3-52*::*URA3*-pRS406
MBY2317	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-T1024Y*
MBY2318	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-R1013H*
MBY2337	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-Y1054F*
MBY2338	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-Y1054A*
MBY2339	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-F1056Y*
MBY2341	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-N1016A*
MBY2342	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-T1024A*
MBY2382	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-Y976A*
MBY2383	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-I972A*
MBY2384	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-Y993A*
MBY2385	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-H1017L*
MBY2386	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-Y967F*
MBY2387	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-G951A*
MBY2395	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-I989A*
MBY2396	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-H1017A*
MBY2397	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-Y1052F*

Table 8. continued.

Name	Genotype
MBY2398	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-Y1052A*
MBY2399	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-Y1052V*
MBY2400	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-E1048A*
MBY2401	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-E1049A*
MBY2402	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-L1050A*
MBY2403	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-R1042H*
MBY2404	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-R1042A*
MBY2405	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-C1068A*
MBY2406	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-C1070A*
MBY2407	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-C1075A*
MBY2408	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-F1056A*
MBY2409	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-E1049G*
MBY2412	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-C1075Y*
MBY2413	MBY2269 *ura3-52*::*URA3*-pRS406-*SET1-H1017R*
MBY2450	*MATa ade2- his3Δ200 ura3-52 ipl1-2 set1Δ::KANMX4*
MBY2469	*MATa ura3-52 trp1- lys2-801 leu2Δ1 his3Δ200 pep4::HIS3 prb1Δ1 can1- GAL BRE2::TAP tag::TRP1#11 set1Δ::KANMX4 #1*
MBY2494	MBY2469 *ura3-52*::*URA3*-pRS406-*SET1*
MBY2496	MBY2469 *ura3-52*::*URA3*-pRS406-*SET1-R1013H*
MBY2511	MBY2450 *ura3-52::URA3*-pRS406-*SET1*
MBY2512	MBY2450 *ura3-52::URA3*-pRS406
MBY2513	MBY2450 *ura3-52::URA3*-pRS406-*SET1-G951A*
MBY2514	MBY2450 *ura3-52::URA3*-pRS406-*SET1-Y967A*
MBY2515	MBY2450 *ura3-52::URA3*-pRS406-*SET1-I989A*
MBY2516	MBY2450 *ura3-52::URA3*-pRS406-*SET1-R1013H*
MBY2517	MBY2450 *ura3-52::URA3*-pRS406-*SET1-N1016A*
MBY2518	MBY2450 *ura3-52::URA3*-pRS406-*SET1-H1017L*
MBY2519	MBY2450 *ura3-52::URA3*-pRS406-*SET1-H1017R*
MBY2520	MBY2450 *ura3-52::URA3*-pRS406-*SET1-C1068A*
MBY2521	MBY2450 *ura3-52::URA3*-pRS406-*SET1-C1075Y*
MBY2522	MBY2450 *ura3-52::URA3*-pRS406-*SET1-Y993A*
MBY2523	MBY2450 *ura3-52::URA3*-pRS406-*SET1-Y1054A*
MBY2524	MBY2450 *ura3-52::URA3*-pRS406-*SET1-T1024A*
MBY2525	MBY2450 *ura3-52::URA3*-pRS406-*SET1-T1024Y*
MBY2526	MBY2450 *ura3-52::URA3*-pRS406-*SET1-R1042A*
MBY2527	MBY2450 *ura3-52::URA3*-pRS406-*SET1-E1048A*
MBY2528	MBY2450 *ura3-52::URA3*-pRS406-*SET1-Y1052V*
MBY2529	MBY2450 *ura3-52::URA3*-pRS406-*SET1-F1056A*
MBY2530	MBY2450 *ura3-52::URA3*-pRS406-*SET1-F1056Y*
MBY2531	MBY2450 *ura3-52::URA3*-pRS406-*SET1-C1070A*
MBY2532	MBY2450 *ura3-52::URA3*-pRS406-*SET1-C1075A*
MBY2540	MBY2469 *ura3-52*::*URA3*-pRS406-*SET1-Y993A*

Table 8. continued.

Name	Genotype
MBY2541	MBY2469 *ura3-52*::*URA3*-pRS406-*SET1-I972A*
MBY2542	MBY2469 *ura3-52*::*URA3*-pRS406-*SET1-G951A*
MBY2543	MBY2469 *ura3-52*::*URA3*-pRS406-*SET1-E1049G*
MBY2544	MBY2469 *ura3-52*::*URA3*-pRS406-*SET1-E1049A*
MBY2546	MBY2469 *ura3-52*::*URA3*-pRS406-*SET1-H1017R*
MBY2547	MBY2469 *ura3-52*::*URA3*-pRS406-*SET1-L1050A*
MBY2548	MBY2469 *ura3-52*::*URA3*-pRS406-*SET1-Y1054F*
MBY2549	MBY2469 *ura3-52*::*URA3*-pRS406-*SET1-Y976F*
MBY2550	MBY2469 *ura3-52*::*URA3*-pRS406-*SET1-R1042H*
MBY2551	MBY2450 *ura3-52::URA3*-pRS406-*SET1-Y967F*
MBY2552	MBY2450 *ura3-52::URA3*-pRS406-*SET1-I972A*
MBY2554	MBY2450 *ura3-52::URA3*-pRS406-*SET1-H1017A*
MBY2556	MBY2450 *ura3-52::URA3*-pRS406-*SET1-R1042H*
MBY2557	MBY2450 *ura3-52::URA3*-pRS406-*SET1-E1049A*
MBY2558	MBY2450 *ura3-52::URA3*-pRS406-*SET1-L1050A*
MBY2559	MBY2450 *ura3-52::URA3*-pRS406-*SET1-Y1052A*
MBY2560	MBY2450 *ura3-52::URA3*-pRS406-*SET1-Y1052F*
MBY2561	MBY2450 *ura3-52::URA3*-pRS406-*SET1-Y1054F*
MBY2563	MBY2450 *ura3-52::URA3*-pRS406-*SET1-E1049G*
MBY2595	MBY2469 *ura3-52*::*URA3*-pRS406-*SET1-H1017A*

Whole cell protein extracts

Cells were grown to ~1-2 X 10^7 cells/ml in 50mL YPADT, harvested and washed with sterile reverse osmosis water. Cells were resuspended in 0.2ml RIPA with 1% SDS and protease inhibitors (150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 50 mM Tris–HCl pH 8.0, 1% SDS, 1 mM phenylmethylsulfonyl fluoride, 1.67 µg/ml aprotonin, 1.67 µg/ml pepstatin, 0.33 µg/ml leupeptin, and 0.40 µg/ml bestatin). Cells were disrupted at 4°C with acid washed glass beads using a Mini-BeadBeater 16 Cell Disrupter (Biospec Products, Bartlesville, OK). Cell lysate was collected and incubated on ice for 30 min, then clarified by centrifugation for 30 min 4°C at 12,000 rpm. Total

protein concentration was determined by a Bradford colorimetric assays using dye reagent concentrate (Bio-Rad, Hercules, CA).

Western blot analysis

All quantitative histone methylation Westerns for Set1 mutant strains in the MBY2269 background were performed by Dr. John Mueller unless otherwise stated. Westerns were carried out as described elsewhere (Mueller *et al.*, 2006) with few modifications. For quantitative histone methylation Westerns, between 8-80 µg of whole cell extract was resolved on 4-20 % Pierce gradient gels (Thermo Fisher Scientific, Rockford, IL) and transferred to polyvinylidene fluoride (PVDF) membranes. Blots were probed with α-K4-monomethyl H3 (07-436, Upstate Cell Signaling or ab8895, Abcam; 1:500), α-K4-dimethyl (07-030, Upstate Cell Signaling; 1:5000), α-K4-trimethyl (ab8580, Abcam; 1:5000), α-histone H3 (ab1791, Abcam; 1:1000), and α-phosphoglycerate kinase (A6457, Molecular Probes: 1:1000). Blots were then probed with either HRP-conjugated α-rabbit or α-mouse secondary antibodies (Promega; 1:2000-1:5000 and 1:2000, respectively), followed by developing with Immuno-Star HRP substrate kit (Bio-Rad, Hercules, CA). Western blots were imaged and quantified using the Molecular Imager ChemiDoc XRS with Quantity One software (Bio-Rad, Hercules, CA).

To measure steady state levels of Set1 protein, ~150 µg of whole cell extract were resolved on 7% SDS-PAGE gels and transferred to PVDF membrane. Blots were probed with α-Set1 antibody (25945, Santa Cruz Biotechnology; 1:1000) and then probed with HRP-conjugated α-goat secondary antibody (2020, Santa Cruz

Biotechnology; 1:1000). Blots were developed and imaged as described above, and then stained with Ponceau S stain to serve as a loading control.

Northern blot analysis

Isolation of total RNA and Northern blotting were performed, as described previously (Bryk et al., 1997; Bryk et al. 2002). Northern blots were imaged with a Pharos FX Plus Molecular Imager and quantified using Quantity One software (Bio-Rad, Hercules, CA).

Nickel purification of His-tagged TEV protease

His tagged Tobacco Etch Virus (TEV) protease was purified using nickel NTA beads (Qiagen, Chatsworth, CA) as described previously (Parks *et al.*, 1995). Following elution from nickel beads, eluate was dialyzed at 4°C for a minimum of 12 hours in 10mM Tris HCl pH 8.0, 0.3M NaCl, and 10% glycerol using a Slide-A-Lyzer Dialysis Cassette (Thermo Fisher Scientific, Rockford, IL).

Tandem affinity purification

Cells were grown to ~5 X 10^7 cells/ml in 6 or 12L of YPADT, harvested and washed with sterile milliQ water. Cell pellet was weighed and resuspended in 1/3 volume of lysis buffer with protease inhibitors (LBpi) forming a cell slurry (40 mM Hepes KOH, 350 mM NaCl, 10% glycerol, 0.1% Tween-20, 1 mM phenylmethylsulfonyl fluoride, 1.67 µg/ml aprotonin, 1.67 µg/ml pepstatin, 0.33 µg/ml leupeptin, and 0.40 µg/ml bestatin). The cell slurry was dripped slowly into liquid N_2.to make yeast popcorn. A prechilled mortar and pestle was used to grind ~10-15g of frozen yeast popcorn into a fine powder, checking the lysis under a microscope (Schultz, 1999;

Wong *et al.*, 1999). Lysed yeast cells were stored at -80°C until needed. For Tandem Affinity Purification (TAP), cell lysate was thawed slowly on ice then purified as previously described (Puig *et al.*, 2001) using 500 µl of Rabbit IgG Agarose beads (Sigma, St. Louis, MO), cleaved with ~100µg of purified His tagged TEV protease, then purified with 500µl of Calmodulin Sepharose 4B beads (GE Healthcare, Pittsburgh, PA). TAP purified COMPASS was eluted from calmodulin beads by incubating for 1 hr at 4°C with a standard elution buffer containing fresh 10 mM EGTA. Eluate was concentrated to ~1-2 µg/µl using Amicon Ultra Centrifugal filters 100K MWCO (Millipore, Billerica, MA).

Plate assay using *ipl1-2 ts* strains

Plate assays were performed by Victoria Schneider. Strains containing *ipl1-2* were grown in 5mL of YPADT and grown to saturation. Ten-fold serial dilutions were made in sterile milliQ water and 5-10 µl of each dilution was spotted onto YPADT agarose plates. Plates were incubated at room temperature for loading and a duplicate at 37°C.

RESULTS

The goal of this study was to design, synthesize, and characterize several mutations in the *SET1* gene that affect histone methylation and silencing at the rDNA. Amino acids were identified using an amino acid alignment of SET domain proteins (Figure 5) as well as the crystal structures of Set7/9 and MLL1. There are five conserved motifs in SET domain proteins: GXG (I), YXG (II), RXXNHS/C (III), EELXY/FXY (IV) and CXCXXXXC (V). Each motif has a specific role in SET domain proteins. GXG (I) motifs are known for their role in binding the cofactor S-adenosyl methionine (SAM) in methyltransferase enzymes (Shields *et al.*, 2003). RXXNHS/C (III) and EELXY/FXY (IV) along with YXG (II) form the active site within a substrate specific channel for the target lysine residue and interact with lysine 4 of histone H3 and SAM (Wilson *et al.*, 2002; Kwon *et al.*, 2003; Xiao *et al.*, 2003). CXCXXXXC (V) coordinates a zinc ion that aides some SET domain proteins in properly folding the c-terminus (Zhang *et al.*, 2003; Southall *et al.*, 2009). Many of the mutants created in yeast Set1 are from these motifs, and are conserved in human Set7/9 and Mll1 as discussed above (Table 2).

Steady-state levels of Set1 and methylated histones *in vi*vo

Figure 5 along with Table 2 details the SET domain mutants created in yeast Set1, focusing on these motifs as well as conserved amino acids. To verify that cells with the wild type or a mutant *SET1* gene integrated at *ura3-52* express Set1 protein, Western blotting assays were performed using whole cell extracts (WCEs) and antibodies specific for *S. cerevisiae* Set1 (Figure 22). The results show that the level of

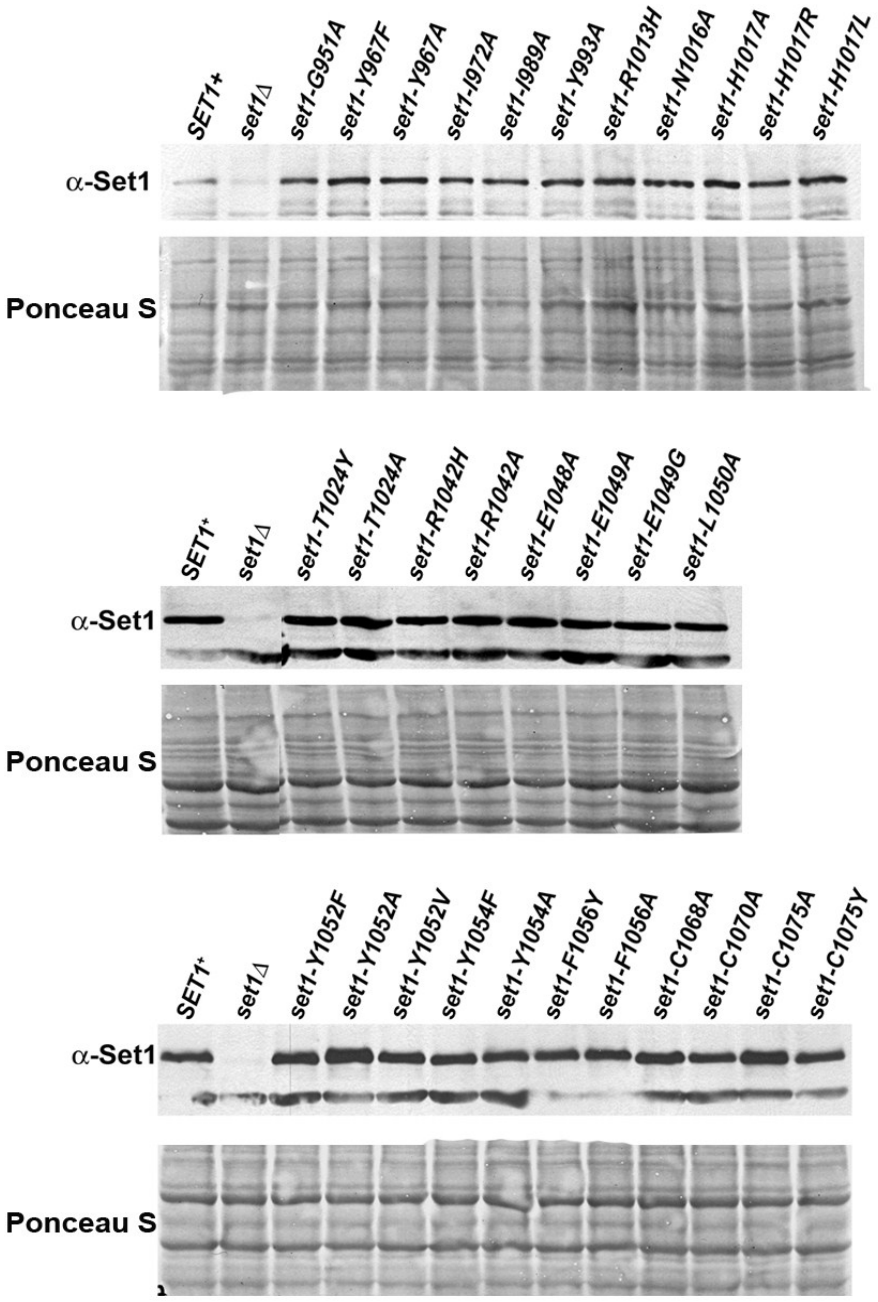


Figure 22. **Western blot analysis of Set1 protein in wild type and mutant strains**. Whole cell extracts (150 μg) were separated by SDS-PAGE and transferred to PVDF membranes. The results of three Western blotting experiments are shown. For each of the three experiments shown, the top panel is an immunoblot showing the level of Set1 protein detected after incubation with anti-Set1 primary antibodies and HRP-conjugated secondary antibodies; the bottom panel is the same membrane shown in the top panel stained with Ponceau S to verify equal loading.

Set1 protein is similar in protein extracts from wild type cells and each of the Set1 mutants. Little or no signal was detected at the position corresponding to Set1 in control extracts from cells that lack Set1 proteins (*set1Δ*).

After verifying that *S. cerevisiae* cells carrying mutations in the *SET1* gene express the Set1 protein at levels similar to those in cells with the wild type *SET1* gene, *in vivo* phenotypes associated with Set1 function were tested. Mutants were evaluated for their ability to methylate H3K4 using Western blots with antibodies specific for each of the three levels of methylated H3K4 with either Pgk1 or total histone H3 protein serving as a loading control (Figure 23, Table 9). By comparing the levels of methylated H3K4 between wild type and mutant WCEs, we were able to separate the Set1 mutants into three classes: mutants with wild type activity, partially active mutants, and null mutants. Mutants similar to or greater than wild type maintained the ability to methylate lysine 4 of histone H3 with levels greater than 50% of wild type. For some of these mutants, levels of methylated H3K4 exceeded the levels detected in wild type cells. The mutants with wild type activity were: I972A, I989A, H1017A, T1024A, T1024Y, R1042H, R1042A, E1048A, L1050A, Y1052F, Y1052V, and F1056Y.

The second class of mutants maintained partial *in vivo* histone H3 methylation activity where the levels of methylated H3K4 were less than 50% of wild type levels. This class of mutants can be further subdivided based on the reduced levels of one or two forms of methylated H3K4. In one subclass of partial H3 methylation mutants containing G951A, Y967F and Y993A, monomethylated H3K4 was detected at a level of 50-70% of wild type, but di- and trimethylated H3K4 were either absent or present at

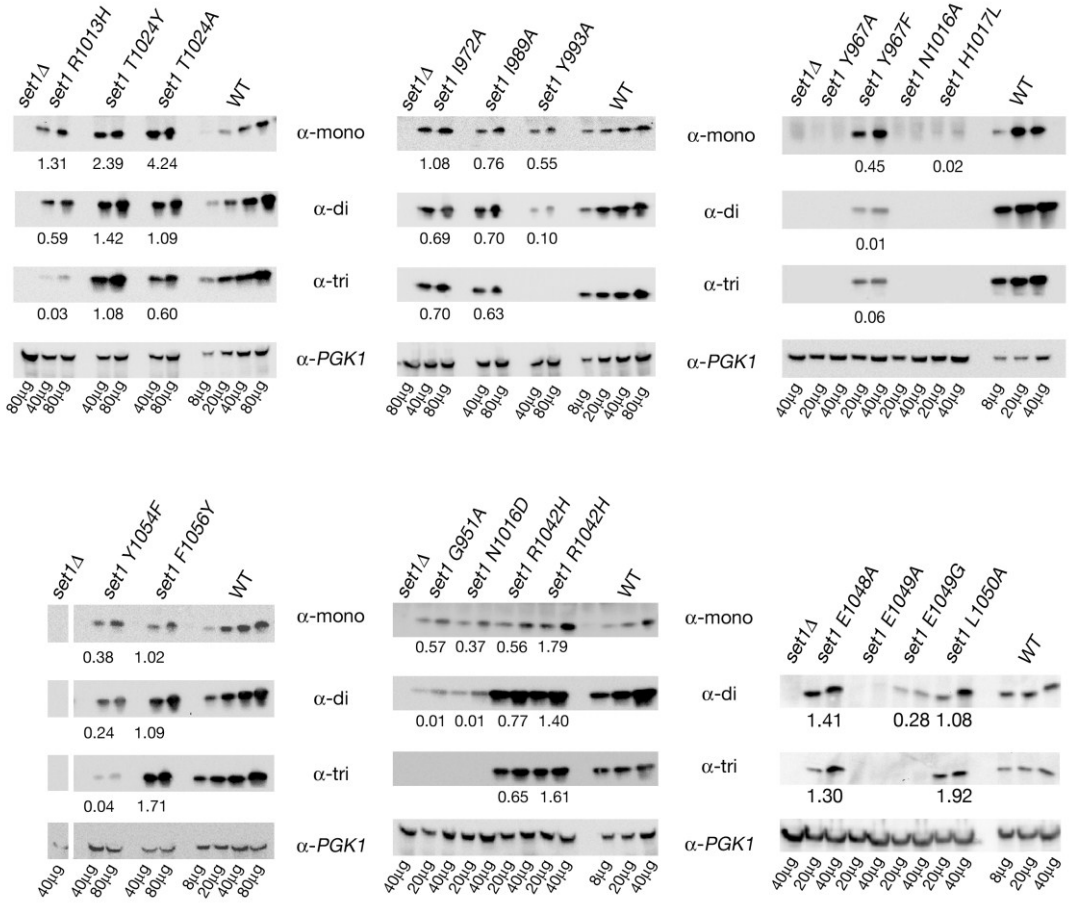


Figure 23. Quantitative Western blots measuring the level of methylated H3K4 in cells expressing wild type and mutant *SET1*. Nine Western blotting experiments are shown. Dilutions of protein extracts (µg loaded is indicated below each loading control panel) from Set1^{+}, *set1Δ* and *set1* mutant cells were analyzed by Western blotting with antibodies to measure the *in vivo* steady-state levels of K4-monomethylated H3 (α-mono), K4-dimethylated H3 (α-di), and K4-trimethylated H3 (α-tri). The level of histone H3 (α-H3) or Pgk1 (α-Pgk1) protein was used to normalize the amount of protein loaded in each lane. The average ratio of normalized K4-methylated H3 detected in *set1* mutant extracts relative to wild type extracts is shown below each blot (n=2 or 3). Results were quantitated using Quantity One software from BioRad. Panels of this figure were generated by John Mueller, with the exception of the EEL domain on the bottom right.

Table 9. Complete data of Quantitative Westerns measuring H3K4 methylation[a].

Mutant	α-monoH3K4		α-diH3K4		α-triH3K4	
	Pgk1/H3	H3	Pgk1/H3	H3	Pgk1/H3	H3
G951A	0.57	0.84	0.01	0.03	0.00	0.00
Y967F	0.45	0.56	0.01	0.01	0.11	0.01
Y967A	0.00	0.00	0.00	0.00	0.00	0.00
I972A	1.08	0.83	0.69	0.99	0.70	0.84
I989A	0.76	0.72	0.70	0.42	0.63	0.63
Y993A	0.55	0.43	0.10	0.02	0.00	0.00
R1013H	1.31	1.35	0.59	0.59	0.03	0.00
N1016A	0.00	0.00	0.00	0.00	0.00	0.00
H1017A	1.65	1.38	1.67	2.28	1.71	4.64
H1017R	0.01	0.04	0.02	0.00	0.00	0.00
H1017L	0.02	0.00	0.00	0.00	0.00	0.00
T1024Y	2.39	4.04	1.42	1.61	1.08	1.14
T1024A	4.24	5.09	1.09	2.39	0.60	1.04
R1042H	0.56*	0.92	0.77*	0.71	0.65*	0.78
R1042A	1.79*	2.09	1.40*	1.24	1.61*	1.34
E1048A	--	--	1.41*	--	1.3*	--
E1049A	--	--	0*	--	0*	--
E1049G	--	--	0.28*	--	0*	--
L1050A	--	--	1.08*	--	1.92*	--
Y1052F	0.54	0.59	0.68	1.00	1.16	1.83
Y1052A	0.19	0.07	0.71	0.48	1.28	0.81
Y1052V	0.83	0.42	1.54	0.93	3.44	1.95
Y1054F	0.38	0.51	0.24	0.32	0.04	0.01
Y1054A	0.00	0.00	0.00	0.00	0.00	0.00
F1056Y	1.02*	0.91	1.09*	0.75	0.41*	0.70
F1056A	0.00	0.00	0.00	0.00	0.00	0.00
C1068A	0	0	0	0	0	0
C1070A	0	0	0	0	0	0
C1075A	0	0	0	0	0	0
C1075Y	0	0	0	0	0	0

[a]The loading control use was either total H3 or Pgk1. "Pgk1/H3" indicates that for those experiments, red values were determined using total H3 as a loading control, while those in black were determined from total Pgk1. Data marked by an asterisk (*) was generated by Kelly M. Williamson, while the rest of the quantitative Westerns were performed by John Mueller.

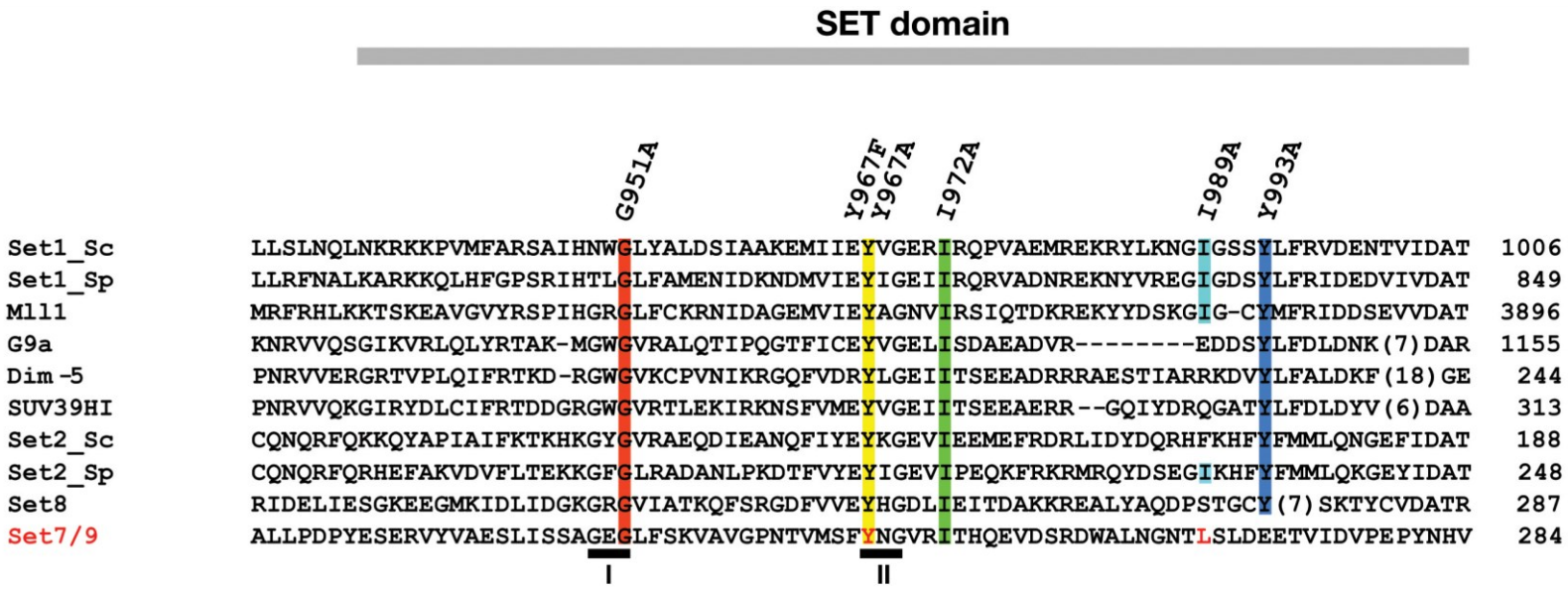
levels that were less than 7% of wild type. In extracts from members of a second subgroup of partial activity mutants (R1013H, E1049G and Y1054F), trimethylated H3K4 was either absent or present at levels less than 5% of wild type. The last subclass of partial activity mutants includes Y1052A that has reduced levels of monomethylated H3K4 but wild type levels of di- and trimethylated H3K4.

The third class, the null mutants, includes mutants that were unable to methylate H3K4 *in vivo*. The null mutants were: Y967A, N1016A, H1017R, H1017L, E1049A, Y1054A, F1056A, C1068A, C1070A, C1075A and C1075Y (Figure 23, Table 9). Set1 protein is expressed at similar levels to wild type in these cells (Figure 22). Therefore, the mutations in each null mutant abolishes methyl transfer activity *in vivo* without affecting the level of expression of the Set1 protein. Table 10 and Figure 24 summarizes these three classes along with additional data discussed below. Further interpretation of the *in vivo* effects of mutations in Set1 will be presented in the discussion and will take into account the results of two additional *in vivo* functions of Set1, Pol II gene silencing at the rDNA (Bryk *et al.*, 2002) and methylation of a Dam1, a kinetochore-associated protein that functions in the spindle checkpoint (Zhang *et al.*, 2005).

Table 10. Summary of characterization of Set1 mutants by methylation of H3K4.

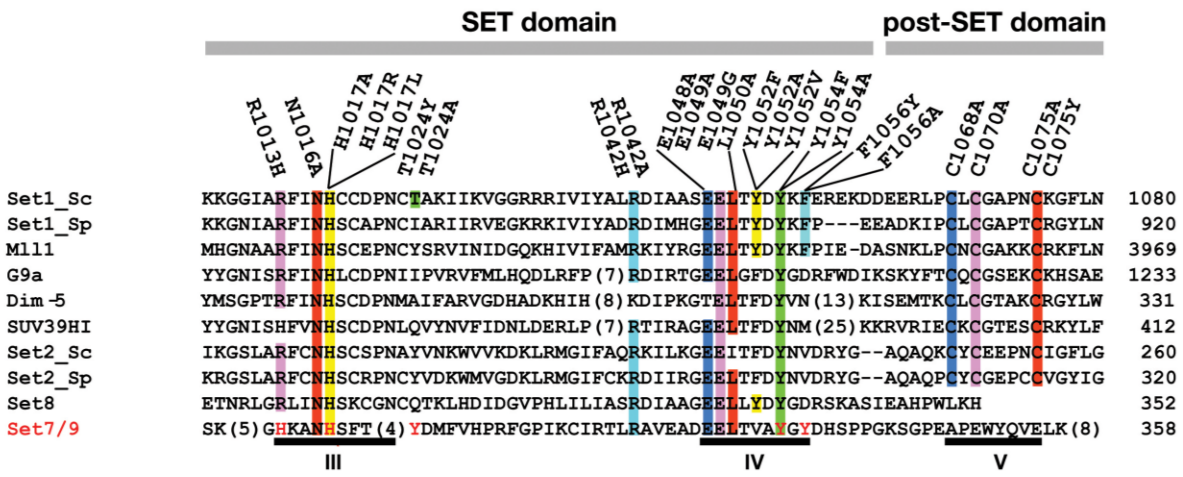
Set1 mutant	Class[a]	rDNA silencing[b]	Suppression of *ipl1-2*[c]
I972A	wild type	proficient	No
I989A	wild type	proficient	No
H1017A	wild type	proficient	No
T1024A	wild type	proficient	No
T1024Y	wild type	proficient	No
R1042A	wild type	proficient	No
R1042H	wild type	defective	No
E1048A	wild type	proficient	No
L1050A	wild type	proficient	No
Y1052F	wild type	proficient	No
Y1052V	wild type	proficient	No
F1056Y	wild type	proficient	No
G951A	partial	defective	partial
Y967F	partial	partial defect	No
Y993A	partial	defective	Yes
R1013H	partial	partial defect	partial
E1049G	partial	defective	partial
Y1054F	partial	defective	partial
Y1052A	partial	proficient	No
Y967A	null	defective	Yes
N1016A	null	defective	Yes
H1017L	null	defective	Yes
H1017R	null	defective	Yes
E1049A	null	defective	Yes
Y1054A	null	defective	Yes
F1056A	null	defective	Yes
C1068A	null	defective	Yes
C1070A	null	defective	Yes
C1075A	null	defective	Yes
C1075Y	null	defective	Yes

[a]Classification based on the levels of K4-methylated H3 detected in whole cell extracts.
[b]Score based on ability to silence Pol II-transcribed gene in rDNA. Lack of rDNA silencing, defective; wild-type level of rDNA silencing, proficient; incomplete rDNA silencing, partial defect.
[c]Ability to suppress *ipl1-2* growth defect at 30°C. Yes, growth at 30°C; no, poor growth at 30°C; partial, intermediate level of growth at 30°C. Plate assays with *ipl1-2* strains were performed by Victoria Schneider.



Mutant	G951A	Y967F	Y967A	I972A	I989A	Y993A
MonoH3K4	0.71	0.51	0.00	0.96	0.74	0.49
DiH3K4	0.02	0.01	0.00	0.84	0.56	0.06
TriH3K4	0.00	0.06	0.00	0.77	0.63	0.00
rDNA silencing	D	PA	D	P	P	D
suppress *ipl1-2*	PA	N	Y	N	N	Y

Figure 24. Alignment of SET domain proteins along with summary of Set1 mutants. This is an alignment of SET domain proteins from several eukaryotes. See Figure 5 for further description of the alignment. Below each section of the alignment is a table summarizing the results of each of the Set1 mutants for H3K4 methylation, rDNA silencing and suppression of *ipl1-2*. The values reported here for mono-, di-, and trimethylated H3K4 are average values of those shown in Table 9. D=defective, P=proficient, PA=partial, Y=Yes, N=No.



Mutant	R1013H	N1016A	H1017A	H1017R	H1017L	T1024Y	T1024A	R1042H	R1042A	E1048A	E1049A	E1049G	L1050A
MonoH3K4	1.33	0.00	1.52	0.03	0.01	3.22	4.67	0.74	1.94	--	--	--	--
DiH3K4	0.59	0.00	1.98	0.01	0.00	1.52	1.74	0.74	1.32	1.41	0.00	0.28	1.08
TriH3K4	0.02	0.00	3.18	0.00	0.00	1.11	0.82	0.71	1.48	1.3	0.00	0.00	1.92
rDNA silencing	PA	D	P	D	D	P	P	D	P	P	D	D	P
suppress *ipl1-2*	PA	Y	N	Y	Y	N	N	N	N	N	Y	PA	N

Mutant	Y1052F	Y1052A	Y1052V	Y1054F	Y1054A	F1056Y	C1068A	C1070A	C1075A	C1075Y
MonoH3K4	0.57	0.13	0.63	0.45	0.00	0.96	0.00	0.00	0.00	0.00
DiH3K4	0.84	0.60	1.24	0.28	0.00	0.92	0.00	0.00	0.00	0.00
TriH3K4	1.50	1.05	2.70	0.03	0.00	0.55	0.00	0.00	0.00	0.00
rDNA silencing	P	P	P	D	D	P	D	D	D	D
suppress *ipl1-2*	N	N	N	PA	Y	N	Y	Y	Y	Y

Figure 24 continued.

Silencing of RNA Pol II transcription at the rDNA

Previous work has shown that Set1 is required for silencing of RNA Pol II transcription at the ribosomal DNA locus (Briggs *et al.*, 2001; Bryk *et al.*, 2002) and at telomeres (Nislow *et al.*, 1997; Krogan *et al.*, 2002). Substitution of lysine 4 of histone H3 with arginine causes a loss of silencing of a RNA Pol II-transcribed gene at the rDNA indicating that methylated H3K4 is required for RNA Pol II gene silencing in *S. cerevisiae* (Briggs *et al.*, 2001; Bryk *et al.*, 2002). Northern blots were performed to determine if mutants of Set1 affected RNA Pol II gene silencing at the rDNA (Figure 22). With the range of H3K4 methylation phenotypes represented in the partial activity Set1 mutants, it was expected that the specific order of methylated H3K4 (mono, di or tri) required for silencing of Pol II genes at the rDNA could be determined. The function of the Set1 mutants in silencing of RNA Pol II transcription was tested using RNA isolated from yeast cells harboring one of the thirty *SET1* mutant alleles and a RNA Pol II-transcribed reporter gene, Ty1*his3AI*, integrated into the rDNA locus (Figure 25). Table 11 reports the average values of n=3-4 Northern blotting experiments and the standard deviation.

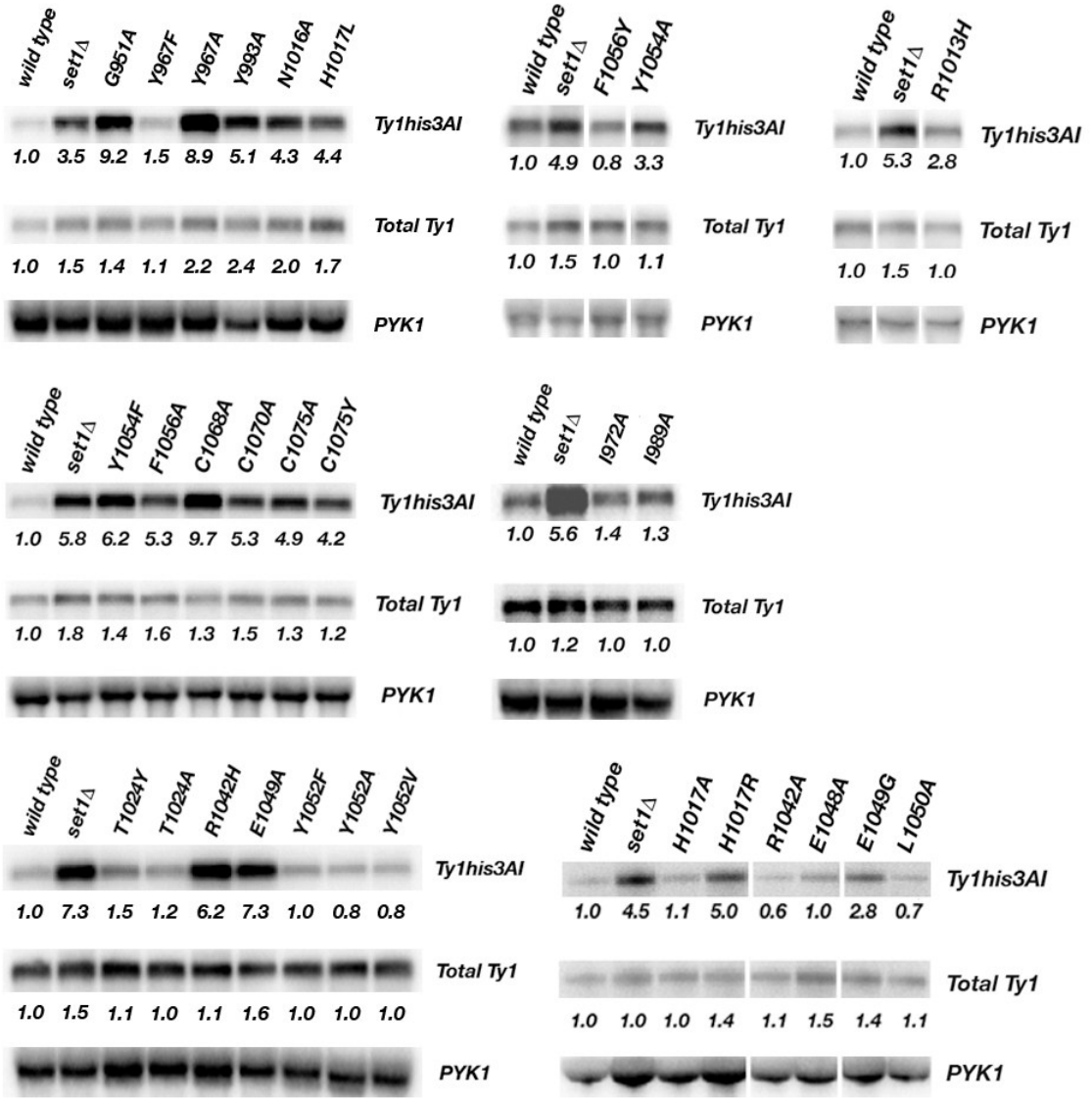


Figure 25. Northern blot analysis to evaluate gene silencing at the rDNA. Seven representative Northern blotting experiments are shown. The sample loaded in each lane is indicated above the top panel of each set of blots. A RNA Pol II-transcribed gene inserted in the rDNA locus, Ty1*his3AI*, was used to assess gene silencing at the rDNA. Total Ty1 mRNA was measured to verify that the effect of the Set1 mutation was specific to the Ty1*his3AI* gene in the rDNA. *PYK1* mRNA was used to normalize the amount of RNA loaded in each lane. The values representing the average ratio of Ty1*his3AI*/*PYK1* mRNA are shown below each Ty1*his3AI* panel, and the values of the average ratio of total Ty1/*PYK1* mRNA are shown below the total Ty1 panels. Each mutant was analyzed in 3 or 4 independent experiments.

Table 11. Complete data of Northern blotting analysis measuring gene silencing at the rDNA[a]

Set1 mutant	HIS3/PYK1	STD	Total Ty1/PYK1	STD
WT	1.00	0.00	1.00	0.00
set1Δ	3.50	0.38	1.51	0.64
G951A	9.19	0.79	1.44	0.62
Y967F	1.53	0.63	1.15	0.69
Y967A	8.92	2.43	2.20	1.14
Y993A	5.09	1.34	2.37	1.01
N1016A	4.26	0.40	1.96	0.92
H1017L	4.37	0.68	1.71	1.14
WT	1.00	0.00	1.00	0.00
set1Δ	4.89	1.81	1.46	0.52
F1056Y	0.84	0.12	0.96	0.28
Y1054A	3.31	0.66	1.09	0.21
WT	1.00	0.00	1.00	0.00
set1Δ	4.86	0.44	1.03	0.42
R1013H	3.10	0.30	1.03	0.10
WT	1.00	0.00	1.00	0.00
set1Δ	5.78	1.60	1.82	0.48
Y1054F	6.24	1.07	1.41	0.15
F1056A	5.35	1.14	1.59	0.49
C1068A	9.67	1.57	1.29	0.26
C1070A	5.28	1.77	1.47	0.56
C1075A	4.88	1.29	1.29	0.26
C1075Y	4.22	1.20	1.23	0.37
WT	1.00	0.00	1.00	0.00
set1Δ	5.62	1.11	1.21	0.07
I972A	1.38	0.51	1.03	0.45
I989A	1.31	0.27	1.03	0.22
WT	1.00	0.00	1.00	0.00
set1Δ	7.28	0.54	1.50	0.11
T1024Y	1.48	0.26	1.06	0.15
T1024A	1.22	0.14	0.99	0.19
R1042H	6.19	0.59	1.14	0.11
E1049A	7.29	1.19	1.56	0.25
Y1052F	1.05	0.01	1.05	0.15
Y1052A	0.77	0.17	1.05	0.08
Y1052V	0.81	0.35	0.99	0.31

Table 11. continued.

Set1 mutant	HIS3/PYK1	STD	Total Ty1/PYK1	STD
WT	1.00	0.00	1.00	0.00
set1Δ	4.50	1.38	1.00	0.38
H1017A	1.10	0.29	1.00	0.34
H1017R	5.00	1.49	1.40	0.51
R1042A	0.60	0.07	1.10	0.20
E1048A	1.00	0.37	1.50	0.89
E1049G	2.80	0.66	1.40	0.73
L1050A	0.70	0.15	1.10	0.38

[a]Data was calculated as a ratio of Ty1*his3AI* to either total Ty1 or PYK1. STD = standard deviation. All Northerns are based on three separate experiments, except for the last set which was repeated four times.

The results of Northern blotting experiments presented in Figure 24 revealed that in most cases the ability of a mutant Set1 protein to maintain silent chromatin at the rDNA correlated directly with its ability to methylate histone H3 *in vivo* (Figure 23). All of the null mutants exhibited a defect in silencing of RNA Pol II transcription consistent with methylated H3K4 being required for the function of silent chromatin at the rDNA. With the exception of R1042H, each mutant with *in vivo* levels of methylated H3K4 that were similar to or greater than wild type retained the ability to silence the Ty1*his3AI* gene in the rDNA. The data largely support the existing model that the requirement for Set1 in gene silencing reflects its role in methylation of H3K4 (Table 10).

The partial activity mutants had the most diverse rDNA-silencing phenotypes and provided insight into the specific form of K4-methylated H3 required for gene silencing at the rDNA. The G951A, Y967F and Y993A mutants share similar defects in methylation of H3 *in vivo*, with moderate levels of monomethylated H3K4 and no di- and trimethylated H3K4 detected in cell extracts. The Y967F mutant exhibited a partial defect while G951A and Y993A each exhibited a complete defect in silencing of the

RNA Pol II reporter gene located in the rDNA. These results suggest that silent chromatin does not function in the presence of moderate levels of monomethylated H3K4 if di- and trimethylated H3K4 are absent. Two other partial activity mutants, Y1052F and Y1052V, with reduced *in vivo* levels of monomethylated H3K4, behaved like wild type cells retaining the ability to silence RNA Pol II gene expression at the rDNA. These results indicate that silent chromatin is maintained in the presence of reduced levels of monomethylated H3K4 when di- and trimethylated H3K4 are present, thereby pointing to a requirement for higher order methylated forms of H3 in the function of silent chromatin in *S. cerevisiae*. Three other partial activity mutants (R1013H, Y1054F and E1049G) with moderate to wild type levels of mono- and dimethylated H3K4, but severely reduced trimethylated H3K4, exhibited defects in gene silencing at the rDNA. Taken together, the results from the partial activity mutants strongly suggest that trimethylated H3K4 is required to silence RNA Pol II transcription at the rDNA. Some of the Set1 mutants produced defects in rDNA silencing greater than *set1Δ*: G951A, Y967A, Y993A, N1016A, H1017L, Y1054F, and C1068A. These mutants may be affecting additional factors that function in silencing at the rDNA although any factors are unknown.

Suppression of *ipl1-2* allele growth defects

In addition to methylating H3K4, Set1 also dimethylates the non-histone protein Dam1 at K233 as part of the anaphase checkpoint (Zhang *et al.*, 2005). Dam1 is part of a large ten subunit complex called the Dam1 complex. Approximately 16-20 Dam1 complexes associate to form a single Dam1 ring which encircles microtubules, sliding

along the microtubule surface creating energy that is used to power chromosome movement and thus segregation of chromosomes during anaphase (Westermann *et al.*, 2007). Repositioning of incorrectly attached microtubules requires the phosphorylation of Dam1 by the *S. cerevisiae* Aurora kinase, Ipl1. Methylation of Dam1 by Set1 prevents phosphorylation of nearby serine residues by Ipl1, a situation that can lead to chromosome loss and growth defects due to missegregation of chromosomes during mitosis. Cells with a point mutation in *IPL1*, *ipl1-2,* have reduced kinase activity and exhibit growth defects due to chromosome loss. Growth defects associated with *ipl1-2* can be suppressed by deletion of Set1, creating a situation where phosphorylation of Dam1 occurs in a manner that not antagonized by Set1, thereby allowing repositioning of incorrectly attached spindle fibers. Because Dam1 and histone H3 are different substrates we expected that the ability to methylate Dam1 would provide an additional means to further separate the Set1 mutants into functional classes.

Cells carrying an *ipl1-2* allele and lacking the endogenous *SET1* gene were transformed with the individual *set1* mutations to evaluate the ability of the Set1 mutants to methylate Dam1 (Materials and Methods). Plate growth assays were performed to evaluate the ability of the Set1 mutants to methylate Dam1 and are summarized in Table 10. The controls, cells with a wild type copy of the *SET1* gene (*ipl1-2 SET1*) and cells lacking SET1 (*ipl1-2 set1Δ*), grew equally well at the permissive temperature (22°C). However, the *ipl1-2 SET1* cells exhibited growth defects at 30°C, a condition that is permissive for the growth of *ipl1-2 set1Δ* cells. As has been shown previously, growth at 30°C occurs in *set1Δ* cells that lack the ability to methylate Dam1 (Zhang *et al.*,

2005). Overall, each of the null mutants suppressed the growth defect of *ipl1-2* allele at 30°C, indicating that methylation of Dam1 does not occur in these mutants. In contrast, none of the mutants with wild type activity were able to suppress the *ipl1-2* growth defect at 30°C suggesting that these Set1 mutant proteins have retained the ability to methylate Dam1.

Purification of COMPASS and *in vitro* analysis

Both Human MLL complexes and Set1A/Set1B complexes contain subunits that are homologues to those found in yeast COMPASS. The MLL family forms multi-subunit complexes characterized by three core subunits of WDR5, RBBP5, and Ash2L forming a core platform to which other MLL proteins can associate (Dou *et al.*, 2006). These subunits have counterparts in the yeast Set1 COMPASS complex through conserved domains: WDR5 is similar to Swd3, RBBP5 is similar to Swd1, and Ash2L is similar to both Bre2 and Spp1. The most studied, WDR5 for WD repeat domain 5, is a doughnut shaped structure with propeller-like domains containing a substrate specific channel through the middle that interacts with dimethylated lysine 4 histone H3 (Wysocka *et al.*, 2005b; Dou *et al.*, 2006). WDR5 is a promiscuous subunit of SET domain methyltransferases and interacts with many methyltransferase complexes like the MLL family and human Set1A/Set1B complex, forming an initial complex of RBBP5 and Ash2L and then recruits MLL1 for methylation (Steward *et al.*, 2006).

Set1A and Set1B associate with the same set of subunits: Ash2L (homolog of Bre2 in yeast), WDR5 (as Swd3 in yeast), RBBP5 (as Swd1 in yeast), CFP1 (as Spp1 in yeast), WDR82/Swd2 (as Swd2 in yeast), and DPY-30 (as Sdc1 in yeast). These

subunits share conserved domains and are thought to act accordingly with similar functions and roles in histone methylation. Both Set1A and Set1B form distinct and separate complexes that each bind to a separate regions of euchromatin, shown by confocal microscopy experiments with differential staining between localization of Set1A and Set1B in human nuclei (Lee *et al.*, 2007). It is believed that human Set1A/Set1B complexes are responsible for ~95% of all methylation, with MLL complexes contributing the remaining 5% of total methylation to specifically targeted HOX genes. Wdr82 is one of the more well studied subunits of human Set1A/Set1B complexes. Wdr82 is a WD40 domain protein similar to WDR5 and a homolog of Swd2 in yeast COMPASS.

Tandem affinity purification (TAP) can be used to evaluate the ability of Set1 mutants to associate with the other subunits of COMPASS as well as using these extracts in *in vitro* assays to investigate the effect these Set1 mutants have on interacting with both SAM and histone H3 through measuring kinetic constants. COMPASS contains eight subunits including Set1, Bre2, Swd1, Spp1, Swd2, Swd3, Sdc1, and Shg1. Previous work from our lab has shown that all subunits, with the exception of Shg1, are required for silencing of RNA Pol II transcription at the rDNA and at telomeres. Deletion of individual subunits of COMPASS results in decreased levels of H3K4 methylation (Mueller *et al.*, 2006). In order to explain Set1 mutants with partial function, TAP was used to investigate whether the association of Set1 is compromised. TAP tagged Bre2 was used to purify COMPASS effectively using first IgG beads followed by cleavage with TEV protease and then a second step of purification with

Calmodulin beads. Proteins from TAP extracts from a wild type strain and two partial function mutants, R1013H and Y967F, were separated using SDS-PAGE gradient gels and silver stained (Figure 26). In each Set1 mutant, seven of the eight subunits of COMPASS can clearly be seen including Set1. Some may argue that Y967F exhibits weakened interaction with Set1. Although the band for Set1 is light, the other six subunits would not be present if Set1 failed to associate with COMPASS. In experiments presented in Roguev *et al*., when purifying COMPASS from a Bre2-TAP strain in a *set1Δ* background only Sdc1 is able to co-purify with Bre2 (Roguev *et al.*, 2001). This is not the case in the TAP results shown in Figure 25.

In vitro assays were performed with purified COMPASS from wild type (lanes 4-9) and R1013H TAP extracts (lanes 13-18) (Figure 27). Results were compared to *in vivo* levels of methylated H3K4 in wild type and *set1Δ* strains. In both wild type and R1013H *in vitro* assays histone methylation can be seen in the absence of added SAM meaning that SAM remains associated with COMPASS throughout an entire purification. Wild type *in vitro* assays show an increase in histone methylation with increasing concentrations of SAM, however, R1013H assays do not suggesting that this mutant has a defect in methylation of histone H3 whether it be due to the mutation affecting binding of H3K4 or due to a solely catalytic defect. In the future, the histone methylation activity of TAP-purified COMPASS will be analyzed *in vitro* in order to further understand the effect these mutants have on the catalytic mechanism of methyl transfer by Set1.

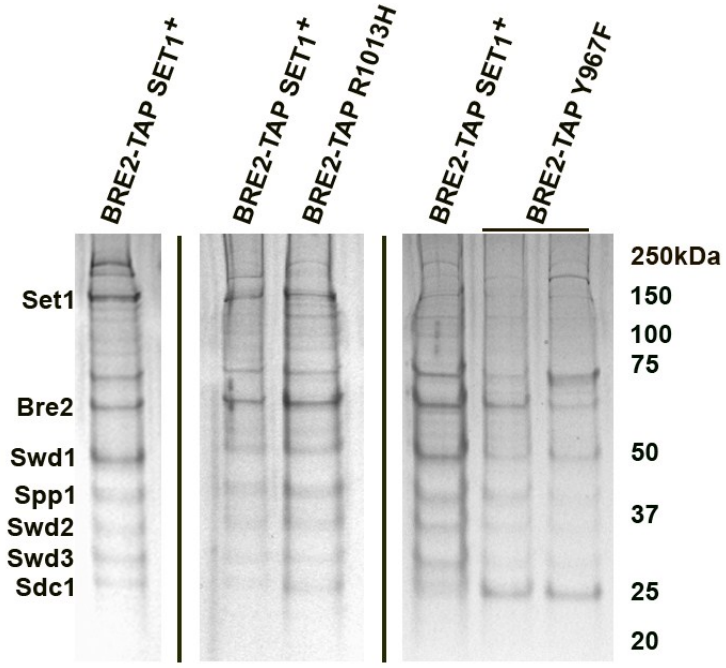


Figure 26. TAP purification of COMPASS from R1013H and Y967F Set1 mutants. COMPASS was purified from cells that express the Bre2 protein fused to the TAP tag in wild type, R1013H, or Y97F strains. Proteins in the kiloDaltons (kDa) ladder and COMPASS were purified by 4-20% SDS-PAGE and visualized by silver staining. COMPASS proteins are indicated to the left and a molecular weight ladder is indicated to the right of the figure. Each of the three panels represents a separate experiment with wild type Set1 Bre2-TAP alongside each mutant.

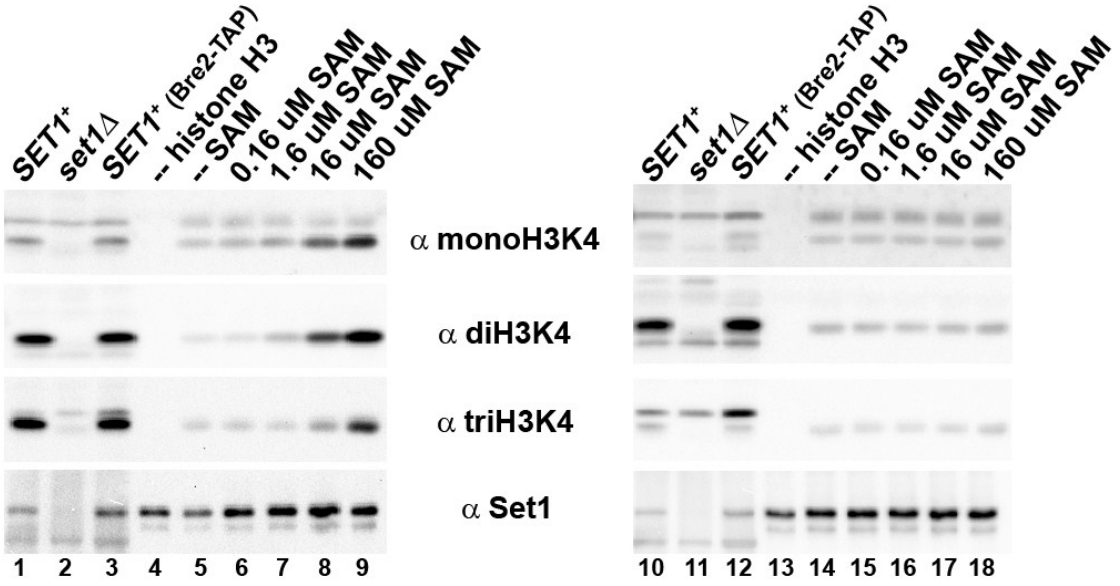


Figure 27. Western analysis of purified COMPASS from wild type and R1013H strains expressing Bre2-TAP. I*n vivo* (lanes 1-3, 10-12) and *in vitro* levels of methylated histone H3 (lanes 4-9, 13-18) were analyzed using Westerns. The left panels contain *in vitro* analysis of wild type purified COMPASS (lanes 4-9) while the right panels contain *in vitro* analysis of R1013H-Set1 purified COMPASS (lanes 13-18). Histone methylation was detected in whole cell extracts of *SET1*$^{+}$ cells (lanes 1 and 3, 10 and 12) but not in extracts from *set1Δ* cells (lane 2, 11).

DISCUSSION

We have designed and characterized thirty mutants within the catalytic SET domain of yeast Set1. Each mutant was characterized for their level of expression *in vivo* with Set1 antisera, for the level of H3K4 methylation *in vivo* with quantitative Westerns, and for their ability to silence an RNA Pol II-transcribed gene inserted into the rDNA through Northerns. R1013H and Y967F were analyzed for their ability to associate with COMPASS using tandem affinity purification, and purified R1013H-COMPASS was analyzed for its ability to methylate full length H3 *in vitro.* Set1 mutants were placed into three categories: wild type mutants, partially active mutants, and null mutants. Several amino acids within the SET domain can be altered without any loss of SET domain activity. In addition, several residues are essential for Set1 activity, producing a complete loss of methylation, yet are still expressed *in vivo* at levels similar to wild type. Nearly all Set1 is thought to incorporate into COMPASS; otherwise free Set1 is unstable (Roguev *et al.*, 2001). The fact that all of the Set1 mutants are expressed and found *in vivo* suggests that each one is incorporated into COMPASS. This makes for an interesting result where a completely catalytically inactive Set1 mutant, such as N1016A, is still able to form COMPASS. In a *set1Δ* strain, it is unknown what other processes may be affected by disassembling the COMPASS complex. In some ways, this mutant would function as a better control than a *set1Δ* strain because it controls for any unknown functions of COMPASS that do not rely on the catalytic ability of Set1. Additional experiments would be needed to determine if a catalytically inactive COMPASS could serve as a control. For example,

ChIP could be used to verify association with RNA Pol II at actively transcribed gene. It would also be important to verify if the proper stoichiometry of COMPASS subunits was maintained in a null mutant.

Trimethylated H3K4 is required for silencing at the rDNA

Our data clearly show that trimethylation of lysine 4 of histone H3 is required for silencing of an RNA Pol II-transcribed gene at the rDNA. The Set1 mutant Y1052A shows reduced levels of monomethylated H3K4 with wild type levels of di- and trimethylated H3K4 and is proficient in rDNA silencing. This mutant suggests that monomethylated H3K4 is not required for silencing at the rDNA. Set1 mutants G951A, Y967F, and Y993A display wild type levels of monomethylated H3K4 with a loss of di- and trimethylated H3K4, and these mutants are all defective for silencing. These results suggest that either di- or trimethylation of H3K4 is required for rDNA silencing. Set1 mutants R1013H and Y1054F show wild type and slightly reduced, respectively, mono- and dimethylated H3K4 with a complete loss of trimethylated H3K4. This data suggests that trimethylated H3K4 is required for silencing at the rDNA.

Set1 mutants classes

Several of the Set1 mutants analyzed produced greater than wild type levels of histone methylation with normal wild type levels of silencing at the rDNA: H1017A, T1024Y, T1024A, R1042A, E1048A, L1050A, Y1052F, and Y1052V. These mutants were also unable to suppress the *ipl1-2* growth defect at 30°C meaning these Set1 mutant proteins are able to properly methylate Dam1. These mutants are prime candidates for further analysis to determine kinetic constants. Some of these Set1 mutants (H1017A,

R1042A, E1048A, and L1050A) result in increases in all three levels of H3K4 methylation *in vivo*. It is likely that these mutants have lost the ability to regulate the rate at which histone methylation occurs relative to wild type; however, only further study will confirm this theory.

Another class of mutants is partial function mutants: G951A, Y967F, Y993A, R1013H, E1049G, Y1052A, and Y1054F. These mutants produce levels of methylated H3K4 less than 50% of wild type levels yet produced conflicting effects on rDNA silencing. For instance, G951A, Y967F, and Y993A all produced ~50% of monomethylated H3K4 with a complete loss of di- and trimethylated H3K4. Yet, G951A and Y993A are defective for silencing while Y967F only gives a slight defect. The same situation exists with R1013H, E1049G, and Y1054F. This group of mutants are targeted for TAP to determine if this conflicting data is a result of subunit composition and for kinetic *in vitro* assays.

GXG motifs are involved in SAM binding

G951A is part of the GXG motif (I) known to be important in binding SAM (Shields *et al.*, 2003). This mutant has relatively wild type levels of monomethylated H3K4 *in vivo*, but a complete loss of di- and trimethylation with a partial defect in Dam1 methylation. Mutation of this glycine to alanine in other methyltransferases observed a decreased ability to bind SAM (Shields *et al.*, 2003). Binding of SAM could be assessed through TAP purification of COMPASS in *in vitro* kinetic assays with increasing amounts of SAM. The higher the amount of SAM required to overcome a binding defect, the weaker the binding is between yeast Set1 and SAM. In addition, in vitro

binding assays could be performed as described in Shields, *et al*. by measuring the amount of bound ^{3}H-SAM to purified G951A-COMPASS using scintillation counting.

Histidine residues in SET domain proteins are responsible for positioning key tyrosine residues

We found that substitutions at His1017 in yeast Set1 are associated with a variety of phenotypes. His1017, along with Arg1013, is part of a conserved motif (III) RXXNHS/C in SET domain proteins (Figure 5). The residues in motif III along with motif II (YXG) and IV (EELXY/FXY) form the active site pocket, with motif III thought to serve a role in binding SAM. Arg1013 and His1017 align with two histidines in Set7/9 theorized to assist Tyr245 and Tyr335 with catalysis by properly positioning these tyrosine residues in the active site (Table 2) (Kwon *et al.*, 2003; Xiao *et al.*, 2003). These residues largely have a structural role and do not make direct contact with either SAM or lysine 4 of histone H3. Positioning of these residues within both human Set7/9 and human MLL1 (Table 2) show the functional group of both residues oriented away from the active site (Figure 7, 9). Although it is unlikely that either yeast Set1 residues (Arg1013 and His1017) form contacts with either SAM or lysine 4 of histone H3, mutations of both residues produced interesting results. Yeast Set1 mutant H1017L produced a null mutant. In addition, H1017R is a null mutant suggesting that the conservation of His1017 in SET motif III is due to more than just the need for a positively charged functional group at that position. Interestingly, H1017A produced a gain of function *SET1* allele, with a 1.5-fold increase of monomethylated H3K4, a 2-fold increase of dimethylated H3K4, and a 3.2-fold increase of trimethylated H3K4. His1017

may serve a more structural role in yeast Set1 although it is unknown if it functions to position any tyrosines within the active site of yeast Set1.

Yeast Set1 Arg1013 aligns with His293 in human Set7/9, and as described above, is one of two histidines in Set7/9 theorized to assist Tyr245 and Tyr335 with catalysis by properly positioning these tyrosine residues in the active site (Kwon *et al.*, 2003; Xiao *et al.*, 2003). R1013H produces a complete loss of trimethylated H3K4 and decreased levels of dimethylated H3K4, in contrast to H1017R which produced a null yeast Set1 mutant. With R1013H, there is a change in the length of the functional group but it is unknown what other amino acids may be affected by this change.

The Phenylalanine/Tyrosine switch in SET domain proteins

In SET domain proteins there is a phenomenon that exists called a phenylalanine/tyrosine switch that controls product specificity. The presence of the hydroxyl group of a tyrosine residue blocks the rotation of methylated lysine 4 of histone H3 and prevents higher levels of methylation (Collins *et al.*, 2005). This switch is used to explain why Set7/9 cannot perform higher levels of methylation and remains a monomethyltransferase. SET domain proteins are rich in tyrosine residues, any one of which could control product specificity. In Set7/9, Tyr245 and Tyr305 are believed to serve this function, while in other SET domain proteins other tyrosine residues may fulfill this role.

Yeast Set1 Tyr967 aligns with Tyr245 in human Set7/9. As mentioned above, Tyr245 is one of two amino acids in Set7/9 studied for its role in product specificity in the phenylalanine/tyrosine switch. Mutant Y245A removes steric hindrance within the

active site by disrupting hydrogen bonding within the active site and allows Set7/9 to trimethylate H3K4 where normally it only monomethylates. The presence of the hydroxyl group of Tyr245 prevents monomethylated H3K4 from rotating and aligning with a new molecule of SAM for further methylation (Xiao *et al.*, 2003; Hu and Zhang, 2006; Zhang and Bruice, 2007). It was thought that mutation of Y967F would allow higher than normal levels of methylation, however the opposite occurred producing a mutant with ~50% of wild type monomethylated H3K4 and a loss of di- and trimethylated H3K4. Furthermore, although the H3K4 methylation profile is nearly identical to G951A and Y993A with a complete loss of silencing at the rDNA, only a slight loss of rDNA silencing is detected (1.5-fold increase over wild type) in Y967F. Y967F is able to methylate Dam1 suggesting that the methylation defect is substrate specific and only affecting methylation of histone H3 and not Dam1. It is possible that substitutions at Y967 weaken interactions with other COMPASS members. Spp1 is a PHD containing protein required for trimethylation of H3K4 and for silencing at the rDNA but is not required for methylation of Dam1 (Schneider *et al.*, 2005; Zhang *et al.*, 2005; Mueller *et al.*, 2006). This could explain why Y967F loses histone methylation but is still able to methylate Dam1. TAP purification of Y967F Set1 (Figure 25) shows seven of the eight subunits can be purified. However, silver staining is not quantitative. Another possibility is that Y967F continues to associate with and methylate H3K4 at silent chromatin but is unable to interact with H3 at active chromatin. This can be tested by ChIP with antisera against Set1 or another COMPASS protein in active and silencing regions of the genome.

Yeast Set1 Tyr993 is part of a previously uncharacterized YXF motif in SET domain proteins and aligns with Tyr3883 of MLL1 (Table 2, Figure 5). The crystal structure of MLL1 shows hydrogen bonds between the adenine ring of SAM and Tyr3883 (Southall *et al.*, 2009). The mutant of Y993A produces 50% of monomethylated H3K4 with a complete loss of di- and trimethylated H3K4. As with G951A, this mutant may affect binding of SAM and will therefore be tested *in vitro* with increasing concentrations of SAM. If there is no effect on SAM binding it is possible Tyr993 is essential for forming hydrogen bonds within the active site.

Thr1024 aligns with Tyr305 of Set7/9 (Table 2), a residue studied for its role in product specificity. The active site contains several water molecules, one of which is close to the amine group of lysine 4 of histone H3 and forms a hydrogen bond with Tyr305. Hydrogen bonding within the active site of Set7/9 is essential in forming a water channel which shuttles protons out of the active site allowing deprotonation of lysine 4 of histone H3, the first step in the mechanism of histone methylation (Wang *et al.*, 2007b; Zhang and Bruice, 2007, 2008a, b). Hydrogen bonding is also necessary for proper alignment of lysine 4 of histone H3 with SAM molecules for methylation to occur (Couture *et al.*, 2006c). Hydrogen bonds between Tyr305, Tyr245, and monomethylated H3K4 hinder the rotation of monomethylated H3K4 and alignment with an additional SAM molecule preventing additional methylation. When mutated to Y305F steric hindrance is removed and allows the monomethyltransferase Set7/9 to dimethylate H3K4 (Kwon *et al.*, 2003; Xiao *et al.*, 2003; Wang *et al.*, 2007b; Zhang and Bruice, 2007; Hu *et al.*, 2008). We expected that removal of the hydroxyl group of

Thr1024 would cause higher levels of trimethylated H3K4 at the expense of mono- and dimethylated H3K4. The presence of a threonine in yeast Set1 could account for why yeast Set1 is able to produce all three levels of histone methylation. Set1 mutants T1024Y and T1024A produce 3- and 4.5-fold increases, respectively, in monomethylated H3K4 *in vivo* with relatively wild type levels of di- and trimethylated H3K4. The presence of this threonine is essential to maintaining wild type levels of histone methylation and it is possible that these mutations affect the rate of histone methylation. It is unlikely that these mutants are functioning by steric hindrance as both a mutation to a tyrosine and alanine residue produce similar results. It is also unlikely that the mutants are affecting hydrogen bonding as T1024Y still retains a hydroxyl group.

Yeast Set1 Tyr1052 aligns with Tyr3942 in human MLL1 (Table 2) and Tyr3942 is considered the Phe/Tyr switch for MLL1. Without additional complex members MLL1 alone is a monomethyltransferase that utilizes positioning of Tyr3942, which is not conserved in Set7/9, to prevent di- and trimethylation (Patel *et al.*, 2009). When MLL1 associates with RBBP5, WDR5, and Ash2L *in vitro*, the complex is still only able to mono- and dimethylate H3K4. Only by mutating Y3942F in an *in vitro* system with MLL1, RBBP5, WDR5, and Ash2L is it able to trimethylate H3K4 (Patel *et al.*, 2009). The orientation of Tyr3942 depends upon whether or not MLL1 binds a histone H3 peptide (Figure 9). Upon binding of a histone H3 peptide, the position of Tyr3942 re-orients and points towards Tyr3858 (Tyr967 in yeast Set1); otherwise, Tyr3942 is positioned ~90° away pointing towards Tyr3914 (Thr1024 in yeast Set1). The

orientation of Tyr3942 shown in Figure 9 is the position it obtains when bound with a histone H3 peptide (Southall *et al.*, 2009). Yeast Set1 is believed to be an exception to the Phe/Tyr switch since it retains a tyrosine residue at 1052 yet can still perform mono-, di-, and trimethylation of H3K4. Y1052F shows a 1.5-fold increase over wild type in trimethylated H3K4, consistent with previously published work who reported that a Y1052F substitution mutant can suppress a trimethylated H3K4 defect associated with deletion of the gene encoding Spp1, a member of the COMPASS complex that stimulates the trimethylation of H3K4 of Set1 (Takahashi *et al.*, 2009). Y1052V produces a 2.7-fold increase over wild type in trimethylated H3K4. Our results show that the hydroxyl group in Y1052 limits trimethylation of H3K4. This restriction if lifted when Tyr1052 is mutated to either phenylalanine or valine but not with alanine. Y1052A is a partial function mutant with reduced monomethylated H3K4 and normal levels of di- and trimethylated H3K4. Overall, the hydroxyl group of Y1052 can limit methylation activity and the size of the amino acid at this position also impacts activity. *In vitro* testing would allow us to measure the rate of each methyl addition and determine how these mutants are affecting catalysis.

Conserved arginine residue involved in protein-protein interactions

Arg1042 is a highly conserved residue among SET domain proteins. In general, arginine residues have a role in protein-protein interactions. Human H3K4 methyltransferases interact with WDR5 (WD repeat domain 5), a doughnut shaped structure with propeller-like domains containing a substrate specific channel through the middle that interacts with dimethylated lysine 4 histone H3 (Wysocka *et al.*, 2005b; Dou

et al., 2006). Crystal structures have clearly shown an N-terminally truncated WDR5 is able to orient and present a tightly bound dimethylated H3K4 substrate to the methyltransferase for trimethylation (Couture *et al.*, 2006b; Han *et al.*, 2006; Ruthenburg *et al.*, 2006). WDR5 is a homolog to yeast Swd proteins and is most similar to Swd2. Co-crystallization of WDR5 with a portion of MLL1 identified an arginine containing domain called *Win* for WDR5 interaction motif that is highly conserved among multicellular organisms (Patel *et al.*, 2008a). *Win* has not been found in yeast. WDR5 recognizes Arg3765 within the C-terminal region of SET domain of MLL1, causing the MLL1 complex to form and become enzymatically active (Patel *et al.*, 2008a; Patel *et al.*, 2008b). This interaction between WDR5 and MLL1 occurs within the same surface as interaction with H3K4 and relies on phenylalanine residues (like Phe133) present on the surface of WDR5 (Song and Kingston, 2008).

Yeast R1042 does not align with Arg3765 in MLL1 and is not located within the same motif (GSARAE) as Arg3765. However, because it is highly conserved there is the possibility that it might function in protein-protein interactions. Our results indicate that both R1042H and R1042A produced wild type levels of H3K4 methylation yet produced different effects for rDNA silencing (Table 10, Figure 24). R1042H is defective for silencing while R1042A is proficient. Previous work from our lab showed that Swd2, the homolog of WDR5, is required for silencing of RNA Pol II transcription at the rDNA (Mueller *et al.*, 2006). If R1042 interacts with Swd2 in yeast, perhaps R1042A is a mutant that is able to bypass the requirement for Swd2 resulting in normal silencing of RNA Pol II transcription at the rDNA while R1042H is unable to bypass the

requirement for Swd2. Further study should include analyzing the subunit composition of COMPASS by TAP purification to determine if any of the Swd proteins in COMPASS are compromised by these mutants.

The EEL domain is important in active site structure

The EEL domain is part of the EELXY/FXY motif that, along with RXXNHS/C and YXG, forms the pseudoknot and overall structure of the active site in SET domain proteins (Jacobs *et al.*, 2002; Kwon *et al.*, 2003; Xiao *et al.*, 2003). E1049 is part of motif IV of the EEL domain that serves a structural role in SET domain proteins and out of the three amino acids E1049 is the most conserved. Yeast Set1 E1049G is a mutant discovered through a random mutagenesis screen as a mutant with partial silencing of RNA Pol II silencing at the rDNA. The data produced supports the idea that E1049 is the critical residue as E1048A and L1050A are wild type in nature (Table 10). Mutation to a smaller residue, either E1049A or E1049G, abolishes H3K4 methylation and the structure of the active site could be compromised. Because the distance between the EEL domain and SAM and lysine 4 of histone H3 it is unlikely that the EEL domain forms direct interactions with either one (Figure 7, 9). Future work will focus on determining the role of E1049 in SET domain proteins, whether it be active site structure or binding of lysine 4 of H3 or SAM. Defects in the structure of the active site could affect overall Set1 structure and thus affect its association with COMPASS. *In vitro* kinetic and binding assays will allow further conclusions on the either ligands.

Mechanism of SET domain proteins

Studies with the crystal structure of Set7/9 have provided two possibilities on how the mechanism of SET domain proteins occurs, primarily the initial step of deprotonation of lysine 4 of histone H3. One model is that an active site base from a tyrosine residue functions to deprotonate lysine 4 while another model suggests the formation of a water channel through hydrogen bonding with tyrosine residues and water molecules in the active site serve as proton acceptors to deprotonate lysine 4 of histone H3. Yeast Set1 Tyr1054 aligns with human Set7/9 Tyr335, an amino acid that has been studied as a possible active site base (Guo and Guo, 2007). Formation of an active site base with Tyr335 has largely been disproven based on computer simulations with Set7/9 that show that while bound with lysine 4 of histone H3 and SAM the pKa of Tyr335 is 18.7 while the pKa of lysine 4 of histone H3 is 8.5 (Zhang and Bruice, 2007). Our data also argues against the presence of an active site base. None of the tyrosine residues mutated to phenylalanine produced a complete loss of histone methylation. Based on the positioning of Tyr335 in Set7/9 (Figure 7) and Tyr3944 in MLL1 (Figure 9) it is likely that Tyr1054 could affect interaction with both lysine 4 of histone H3 and SAM. Y1054F is a partial function mutant with reduced mono- and dimethylated H3K4 and a loss of trimethylated H3K4 *in vivo*. The hydroxyl group of Tyr1054 is required for full levels of methylation but the presence of the phenyl ring, by Y1054F, still permits low levels of methylation. Y1054A shows a complete loss of all methylation. Future work will focus on an *in vitro* analysis to further analyze the effect these mutants have on catalysis.

CHAPTER IV

SUMMARY AND CONCLUSIONS

My research has focused on understanding how chromatin transitions from actively transcribed euchromatin to silent heterochromatin and factors that affect this transition, namely Set1 a protein methyltransferase that targets lysine 4 of histone H3 and Dam1. This change in chromatin forms some of our fundamental questions of how chromatin structure and protein methylation affects gene expression. One of my projects consisted of using a single molecule technique to analyze changes in chromatin structure through the in vivo expression of a DNA methyltransferase *M.CviPI*. My second project focused on identifying critical residues in yeast Set1involved in binding of SAM and lysine 4 of histone H3, in the phenylalanine/tyrosine switch, and in the mechanism of SET domain proteins.

In vivo* expression of *M.CviPI

Chapter II details a single molecule technique that was used to analyze changes in chromatin structure through the *in vivo* expression of a DNA methyltransferase *M.CviPI.* These experiments focused on the NTS2 region in the rDNA, a region our lab has previously studied using MNase mapping experiments (Li *et al.*, 2006a). With an in vivo expression of *M.CviPI*, we hoped to gain knowledge of the changes in chromatin structure within NTS2 region of the rDNA due to a silencing defect caused by the deletion of *SIR2* and due to RNA Pol I transcription by using a 25 rDNA copy array with all twenty five 35S rRNA genes actively transcribed by RNA Pol I. Our experiments

were based on the hypothesis that an actively transcribed 35S rRNA gene remains active throughout a single cell cycle. Overall, we observed changes in positioning of nucleosomes 5, 4, and 3 in *sir2Δ* and 25 copy strains. Furthermore, we have shown the presence of a bound factor within the 35S rRNA promoter that associates exclusively with actively transcribed 35S rRNA genes within the upstream element which is likely to be Hmo1.

Although we have been able to utilize a single molecule technique to observe changes in nucleosome positioning there are various limitations to using this technique. Few conclusions can be drawn from visualizing single molecules of our data, like those displayed in Figure 15 due to the sheer number of clones produced in each experiment. The data must be summarized and visualized as a whole in order to obtain information on how the chromatin structure is changing. In order to visualize nucleosome positioning, sequencing data is summarized in graphs such as Figure 13. The correlation analyses combine and summarize the data as a whole in order to show changes in chromatin structure (Figure 14). In addition, although we have focused on a single region of the rDNA our results are still obtained from a population of cells in various stages of the cell cycle.

Data presented by Tan, R.Z. and Van Oudenaarden, *et al.* 2010 changes the outlook of this project. A 35S rRNA gene is only transcribed for five minutes by RNA Pol I before becoming inactive; however, we induced *M.CviPI* for far longer. Single molecules, like that in Figure 15, do not represent open and closed repeats. Instead, our single molecules likely represent a mixture of open and closed repeats. Because of this,

single molecules will not be easily segregated in to open and closed classes because they will not contain definable features that will discriminate between open and closed chromatin. Future work will need to be done to find a way of analyzing this data and producing a significance value. Dr. Mary Bryk has been working with Dr. Heather Wilkinson from the Department of Plant Pathology and Microbiology at Texas A&M University on new ways of analyzing our data. One way is through utilizing an outlier distance plot which compares an unknown data set to a known data set. Since the 25 copy array is actively transcribed by RNA Pol I and has an open chromatin structure, this serves as a basis of comparison in other strains to determine differences in accessibility and chromatin structure. Included in Dr. Wilkinson's work is forming a dendrogram to show how different data sets cluster in relation to each other based on accessibility to *M.CviPI*. This is another technique that will be used in the future to separate clones into classes based on accessibility to *M.CviPI*.

With that said, the project is significant in that it will lay the foundation to understand the physical structure of actively transcribed rDNA repeats and silent repeats using a single molecule approach. It has long been known that the rDNA contains half of the repeats actively transcribed by RNA Pol I and the other half silent. To be able to definitively say what characterizes the structure of active and inactive rDNA repeats using a single molecule technique would provide much more detail than a technique that gives average changes, such as MNase and psoralen. The project has the possibility to provide the lab with a new tool to analyze the rDNA that can easily be assessed with statistical methods such as the correlation analyses presented in Chapter II.

Mutagenesis of yeast Set1

Chapter III describes a second area of research which focused on a mutagenesis of yeast Set1. We have focused on critical residues in yeast Set1involved in binding of SAM and lysine 4 of histone H3, in the phenylalanine/tyrosine switch, and in the mechanism of SET domain proteins. Thirty mutants were created and characterized for their effect on *in vivo* H3K4 methylation, silencing of RNA Pol II transcription at the rDNA, and methylation of Dam1. Y967F and R1013H Set1 mutants were purified from whole cell extracts to verify their association with additional complex members of COMPASS. *In vitro* assays of wild type and R1013H COMPASS were analyzed by Western blot, showing a defect in H3K4 methylation for R1013H. Our mutants have shown that trimethylated H3K4 is required for silencing at the rDNA. These mutants also provide information on the presence of a phenylalanine/tyrosine switch in SET domain proteins by mutating several tyrosine residues to phenylalanine and to alanine. Several of these conserved tyrosine residues produced interesting results. Y967F and Y993A produced a mutant with only monomethylated H3K4. T1024Y/A produced mutants with higher than normal monomethylated H3K4 suggesting that these mutants have lost the ability to regulate H3K4 methylation.

These mutants have also provided a greater understanding on the catalytic mechanism of SET domain proteins, arguing against the presence of a base-catalyzed reaction and towards the use of hydrogen bonding and formation of a water channel in yeast Set1 for the deprotonation of lysine 4 of histone H3. None of the tyrosine to phenylalanine mutants would have produced a complete loss of H3K4 methylation.

Conclusions about these mutants are limited because *in vitro* kinetic and binding assays have not been performed. Many of the yeast Set1 mutants discussed above would be prime candidates for these *in vitro* assays. There are several ways this can be accomplished. One way would be to use i*n vitro* assays with ^{3}H-SAM in filter binding assays modeled after an article published on using such assays to study histone acetyltransferases (Berndsen and Denu, 2005). This article describes in detail how to perform such assays and analyze the results to obtain Km, Vmax, kcat, and V/K. These experiments would be carried out with recombinant full length H3 purified from inclusion bodies from *E. coli*. A drawback of these experiments is that it would only provide an overall rate for the conversion of unmodified H3 to methylated H3. This could possibly be addressed by first treating full length H3 with Set7/9, a monomethyltransferase purchased from New England Biolabs (NEB). This would create a second substrate to use in these assays, monomethylated H3K4.

A second avenue would be to use histone H3 peptides purchased as unmodified, monomethylated and dimethylated H3K4, and use these peptides in assays to determine each individual rate. It is unknown if small peptides could be used in these filter binding assays as conflicting reports of using peptides have been found in the literature. If not, *in vitro* assays could be analyzed using mass spectroscopy to identify the products of these reactions. Dr. Michelle Henderson at Texas A&M University has been working with Dr. Larry Dangott in using mass spectroscopy to quantitatively analyze products produced in *in vitro* assays. *In vitro* binding assays could be modeled after a paper studying GXG motifs and their role in binding SAM (Shields *et al.*, 2003). As above,

^{3}H-SAM would be utilized with purified mutant COMPASS followed by measuring bound and unbound ^{3}H-SAM through scintillation counting.

Results from kinetic assays would be relatively new to the field where others who have attempted these experiments have met with little success and questionable results. Even if using peptides is not feasible, using full length H3 and monomethylated H3, as mentioned above, would still allow us the ability to measure the conversion to monomethylated H3K4 using unmodified H3, and to measure the conversion to trimethylated H3K4 using monomethylated H3. Again, these results would provide information about how these mutants are affecting the ability of Set1 to methylated lysine 4 of histone H3. Binding assays discussed above would also provide ample information about how these Set1 mutants are affecting association of SAM and lysine 4 of histone H3. Conclusions about these mutants would enable us to provide explanations on which residues are critical for interactions with SAM and lysine 4 of histone H3, on how these mutants affect the mechanism of histone methylation, and which mutants are critical for that reaction to occur. Beyond *in vitro* assays, these mutants also provide the possibility of understanding how association of the subunits of COMPASS might be affected. This would result in being able to draw conclusions about the overall COMPASS structure and composition, something that is lacking in our field. Overall, our results have a greater implication in understanding human H3K4 methyltransferases and their role in childhood leukemia and brain cancer, colon cancer, and diabetes.

REFERENCES

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc Natl Acad Sci USA *51*, 786-794.

Allfrey, V.G., and Mirsky, A.E. (1964). Structural modifications of histones and their possible role in the regulation of RNA Synthesis. Science *144*, 559.

Andreu-Vieyra, C.V., Chen, R., Agno, J.E., Glaser, S., Anastassiadis, K., Stewart, A.F., and Matzuk, M.M. (2010). MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. PLoS Biol *8*.

Andrulis, E.D., Neiman, A.M., Zappulla, D.C., and Sternglanz, R. (1998). Perinuclear localization of chromatin facilitates transcriptional silencing. Nature *394*, 592-595.

Andrulis, E.D., Zappulla, D.C., Ansari, A., Perrod, S., Laiosa, C.V., Gartenberg, M.R., and Sternglanz, R. (2002). Esc1, a nuclear periphery protein required for Sir4-based plasmid anchoring and partitioning. Mol Cell Biol *22*, 8292-8301.

Aparicio, O.M., Billington, B.L., and Gottschling, D.E. (1991). Modifiers of position effect are shared between telomeric and silent mating-type loci in *S. cerevisiae*. Cell *66*, 1279-1287.

Aprikian, P., Moorefield, B., and Reeder, R.H. (2001). New model for the yeast RNA polymerase I transcription cycle. Mol Cell Biol *21*, 4847-4855.

Ayton, P.M., Chen, E.H., and Cleary, M.L. (2004). Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. Mol Cell Biol *24*, 10470-10478.

Azzam, R., Chen, S.L., Shou, W., Mah, A.S., Alexandru, G., Nasmyth, K., Annan, R.S., Carr, S.A., and Deshaies, R.J. (2004). Phosphorylation by cyclin B-Cdk underlies release of mitotic exit activator Cdc14 from the nucleolus. Science *305*, 516-519.

Bach, C., Mueller, D., Buhl, S., Garcia-Cuellar, M.P., and Slany, R.K. (2009). Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2. Oncogene *28*, 815-823.

Bannister, A.J., Schneider, R., and Kouzarides, T. (2002). Histone methylation: dynamic or static? Cell *109*, 801-806.

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., and de Massy, B. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science *327*, 836-840.

Bazett-Jones, D.P., Leblanc, B., Herfort, M., and Moss, T. (1994). Short-range DNA looping by the Xenopus HMG-box transcription factor, xUBF. Science *264*, 1134-1137.

Beisel, C., Imhof, A., Greene, J., Kremmer, E., and Sauer, F. (2002). Histone methylation by the Drosophila epigenetic transcriptional regulator Ash1. Nature *419*, 857-862.

Berger, S.L. (2002). Histone modifications in transcriptional regulation. Curr Opin Genet Dev *12*, 142-148.

Berndsen, C.E., and Denu, J.M. (2005). Assays for mechanistic investigations of protein/histone acetyltransferases. Methods *36*, 321-331.

Bernstein, H.J. (2000). Recent changes to RasMol, recombining the variants. Trends Biochem Sci *25*, 453-455.

Bier, M., Fath, S., and Tschochner, H. (2004). The composition of the RNA polymerase I transcription machinery switches from initiation to elongation mode. FEBS Lett *564*, 41-46.

Birke, M., Schreiner, S., Garcia-Cuellar, M.P., Mahr, K., Titgemeyer, F., and Slany, R.K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. Nucleic Acids Res *30*, 958-965.

Blobel, G.A., Kadauke, S., Wang, E., Lau, A.W., Zuber, J., Chou, M.M., and Vakoc, C.R. (2009). A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. Mol Cell *36*, 970-983.

Bose, M.E., McConnell, K.H., Gardner-Aukema, K.A., Muller, U., Weinreich, M., Keck, J.L., and Fox, C.A. (2004). The origin recognition complex and Sir4 protein recruit Sir1p to yeast silent chromatin through independent interactions requiring a common Sir1p domain. Mol Cell Biol *24*, 774-786.

Boulton, S.J., and Jackson, S.P. (1998). Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. EMBO J *17*, 1819-1828.

Braunstein, M., Rose, A.B., Holmes, S.G., Allis, C.D., and Broach, J.R. (1993). Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. Genes Dev *7*, 592-604.

Braunstein, M., Sobel, R.E., Allis, C.D., Turner, B.M., and Broach, J.R. (1996). Efficient transcriptional silencing in *Saccharomyces cerevisiae* requires a heterochromatin histone acetylation pattern. Mol Cell Biol *16*, 4349-4356.

Brewer, B.J., and Fangman, W.L. (1987). The localization of replication origins on ARS plasmids in *S. cerevisiae*. Cell *51*, 463-471.

Brewer, B.J., and Fangman, W.L. (1988). A replication fork barrier at the 3' end of yeast ribosomal RNA genes. Cell *55*, 637-643.

Brewer, B.J., and Fangman, W.L. (1994). Initiation preference at a yeast origin of replication. Proc Natl Acad Sci U S A *91*, 3418-3422.

Brewer, B.J., Lockshon, D., and Fangman, W.L. (1992). The arrest of replication forks in the rDNA of yeast occurs independently of transcription. Cell *71*, 267-276.

Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F., and Allis, C.D. (2001). Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. Genes Dev *15*, 3286-3295.

Bryk, M., Banerjee, M., Murphy, M., Knudsen, K.E., Garfinkel, D.J., and Curcio, M.J. (1997). Transcriptional silencing of Ty1 elements in the RDN1 locus of yeast. Genes Dev *11*, 255-269.

Bryk, M., Briggs, S.D., Strahl, B.D., Curcio, M.J., Allis, C.D., and Winston, F. (2002). Evidence that Set1, a factor required for methylation of histone H3, regulates rDNA silencing in *S. cerevisiae* by a Sir2-independent mechanism. Curr Biol *12*, 165-170.

Buck, S.W., Sandmeier, J.J., and Smith, J.S. (2002). RNA polymerase I propagates unidirectional spreading of rDNA silent chromatin. Cell *111*, 1003-1014.

Carvin, C.D., Dhasarathy, A., Friesenhahn, L.B., Jessen, W.J., and Kladde, M.P. (2003). Targeted cytosine methylation for in vivo detection of protein-DNA interactions. Proc Natl Acad Sci U S A *100*, 7743-7748.

Caslini, C., Alarcon, A.S., Hess, J.L., Tanaka, R., Murti, K.G., and Biondi, A. (2000). The amino terminus targets the mixed lineage leukemia (MLL) protein to the nucleolus, nuclear matrix and mitotic chromosomal scaffolds. Leukemia *14*, 1898-1908.

Caslini, C., Connelly, J.A., Serna, A., Broccoli, D., and Hess, J.L. (2009). MLL associates with telomeres and regulates telomeric repeat-containing RNA transcription. Mol Cell Biol *29*, 4519-4526.

Chang, P.Y., Hom, R.A., Musselman, C.A., Zhu, L., Kuo, A., Gozani, O., Kutateladze, T.G., and Cleary, M.L. (2010). Binding of the MLL PHD3 finger to histone H3K4me3 is required for MLL-dependent gene transcription. J Mol Biol *400*, 137-144.

Cheeseman, I.M., Anderson, S., Jwa, M., Green, E.M., Kang, J., Yates, J.R., 3rd, Chan, C.S., Drubin, D.G., and Barnes, G. (2002). Phospho-regulation of kinetochore-microtubule attachments by the Aurora kinase Ipl1p. Cell *111*, 163-172.

Cheeseman, I.M., Enquist-Newman, M., Muller-Reichert, T., Drubin, D.G., and Barnes, G. (2001). Mitotic spindle integrity and kinetochore function linked by the Duo1p/Dam1p complex. J Cell Biol *152*, 197-212.

Chen, J., Santillan, D.A., Koonce, M., Wei, W., Luo, R., Thirman, M.J., Zeleznik-Le, N.J., and Diaz, M.O. (2008). Loss of MLL PHD finger 3 is necessary for MLL-ENL-induced hematopoietic stem cell immortalization. Cancer Res *68*, 6199-6207.

Cheng, F., Liu, J., Zhou, S.H., Wang, X.N., Chew, J.F., and Deng, L.W. (2008). RNA interference against mixed lineage leukemia 5 resulted in cell cycle arrest. Int J Biochem Cell Biol *40*, 2472-2481.

Chiang, P.K., Gordon, R.K., Tal, J., Zeng, G.C., Doctor, B.P., Pardhasaradhi, K., and McCann, P.P. (1996). S-Adenosylmethionine and methylation. FASEB J *10*, 471-480.

Choe, S.Y., Schultz, M.C., and Reeder, R.H. (1992). In vitro definition of the yeast RNA polymerase I promoter. Nucleic Acids Res *20*, 279-285.

Chuikov, S., Kurash, J.K., Wilson, J.R., Xiao, B., Justin, N., Ivanov, G.S., McKinney, K., Tempst, P., Prives, C., Gamblin, S.J., Barlev, N.A., and Reinberg, D. (2004). Regulation of p53 activity through lysine methylation. Nature *432*, 353-360.

Cioci, F., Vogelauer, M., and Camilloni, G. (2002). Acetylation and accessibility of rDNA chromatin in *Saccharomyces cerevisiae* in (Delta)top1 and (Delta)sir2 mutants. J Mol Biol *322*, 41-52.

Cioci, F., Vu, L., Eliason, K., Oakes, M., Siddiqi, I.N., and Nomura, M. (2003). Silencing in yeast rDNA chromatin: reciprocal relationship in gene expression between RNA polymerase I and II. Mol Cell *12*, 135-145.

Claypool, J.A., French, S.L., Johzuka, K., Eliason, K., Vu, L., Dodd, J.A., Beyer, A.L., and Nomura, M. (2004). Tor pathway regulates Rrn3p-dependent recruitment of yeast RNA polymerase I to the promoter but does not participate in alteration of the number of active genes. Mol Biol Cell *15*, 946-956.

Clemente-Blanco, A., Mayan-Santos, M., Schneider, D.A., Machin, F., Jarmuz, A., Tschochner, H., and Aragon, L. (2009). Cdc14 inhibits transcription by RNA polymerase I during anaphase. Nature *458*, 219-222.

Collins, R.E., Tachibana, M., Tamaru, H., Smith, K.M., Jia, D., Zhang, X., Selker, E.U., Shinkai, Y., and Cheng, X. (2005). In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. J Biol Chem *280*, 5563-5570.

Couture, J.F., Collazo, E., Hauk, G., and Trievel, R.C. (2006a). Structural basis for the methylation site specificity of SET7/9. Nat Struct Mol Biol *13*, 140-146.

Couture, J.F., Collazo, E., and Trievel, R.C. (2006b). Molecular recognition of histone H3 by the WD40 protein WDR5. Nat Struct Mol Biol *13*, 698-703.

Couture, J.F., Hauk, G., Thompson, M.J., Blackburn, G.M., and Trievel, R.C. (2006c). Catalytic roles for carbon-oxygen hydrogen bonding in SET domain lysine methyltransferases. J Biol Chem *281*, 19280-19287.

D'Ambrosio, C., Schmidt, C.K., Katou, Y., Kelly, G., Itoh, T., Shirahige, K., and Uhlmann, F. (2008). Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. Genes Dev *22*, 2215-2227.

Dammann, R., Lucchini, R., Koller, T., and Sogo, J.M. (1993). Chromatin structures and transcription of rDNA in yeast *Saccharomyces cerevisiae*. Nucleic Acids Res *21*, 2331-2338.

Dammann, R., Lucchini, R., Koller, T., and Sogo, J.M. (1995). Transcription in the yeast rRNA gene locus: distribution of the active gene copies and chromatin structure of their flanking regulatory sequences. Mol Cell Biol *15*, 5294-5303.

D'Amours, D., Stegmeier, F., and Amon, A. (2004). Cdc14 and condensin control the dissolution of cohesin-independent chromosome linkages at repeated DNA. Cell *117*, 455-469.

de Bruin, D., Kantrow, S.M., Liberatore, R.A., and Zakian, V.A. (2000). Telomere folding is required for the stable maintenance of telomere position effects in yeast. Mol Cell Biol *20*, 7991-8000.

Deering, T.G., Ogihara, T., Trace, A.P., Maier, B., and Mirmira, R.G. (2009). Methyltransferase Set7/9 maintains transcription and euchromatin structure at islet-enriched genes. Diabetes *58*, 185-193.

Deininger, P.L., and Batzer, M.A. (1999). Alu repeats and human disease. Mol Genet Metab *67*, 183-193.

DeLange, R.J., Fambrough, D.M., Smith, E.L., and Bonner, J. (1969). Calf and pea histone IV. II. The complete amino acid sequence of calf thymus histone IV; presence of epsilon-N-acetyllysine. J Biol Chem *244*, 319-334.

Del Rizzo, P.A., Couture, J.F., Dirk, L.M., Strunk, B.S., Roiko, M.S., Brunzelle, J.S., Houtz, R.L., and Trievel, R.C. (2010). SET7/9 catalytic mutants reveal the role of active site water molecules in lysine multiple methylation. J Biol Chem *285*, 31849-31858.

Deng, L.W., Chiu, I., and Strominger, J.L. (2004). MLL 5 protein forms intranuclear foci, and overexpression inhibits cell cycle progression. Proc Natl Acad Sci USA *101*, 757-762.

Devine, S.E., and Boeke, J.D. (1996). Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes Dev *10*, 620-633.

Dirk, L.M., Flynn, E.M., Dietzel, K., Couture, J.F., Trievel, R.C., and Houtz, R.L. (2007). Kinetic manifestation of processivity during multiple methylations catalyzed by SET domain protein methyltransferases. Biochemistry *46*, 3905-3915.

Djabali, M., Selleri, L., Parry, P., Bower, M., Young, B.D., and Evans, G.A. (1992). A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. Nat Genet *2*, 113-118.

Dou, Y., Milne, T.A., Ruthenburg, A.J., Lee, S., Lee, J.W., Verdine, G.L., Allis, C.D., and Roeder, R.G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat Struct Mol Biol *13*, 713-719.

Dover, J., Schneider, J., Tawiah-Boateng, M.A., Wood, A., Dean, K., Johnston, M., and Shilatifard, A. (2002). Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. J Biol Chem *277*, 28368-28371.

Emerling, B.M., Bonifas, J., Kratz, C.P., Donovan, S., Taylor, B.R., Green, E.D., Le Beau, M.M., and Shannon, K.M. (2002). MLL5, a homolog of Drosophila trithorax located within a segment of chromosome band 7q22 implicated in myeloid leukemia. Oncogene *21*, 4849-4854.

Enquist-Newman, M., Cheeseman, I.M., Van Goor, D., Drubin, D.G., Meluh, P.B., and Barnes, G. (2001). Dad1p, third component of the Duo1p/Dam1p complex involved in kinetochore function and mitotic spindle integrity. Mol Biol Cell *12*, 2601-2613.

Fair, K., Anderson, M., Bulanova, E., Mi, H., Tropschug, M., and Diaz, M.O. (2001). Protein interactions of the MLL PHD fingers modulate MLL target gene regulation in human cells. Mol Cell Biol *21*, 3589-3597.

Fang, J., Hogan, G.J., Liang, G., Lieb, J.D., and Zhang, Y. (2007). The *Saccharomyces cerevisiae* histone demethylase Jhd1 fine-tunes the distribution of H3K36me2. Mol Cell Biol *27*, 5055-5065.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448-453.

Ferreira, J., and Carmo-Fonseca, M. (1997). Genome replication in early mouse embryos follows a defined temporal and spatial order. J Cell Sci *110 (Pt 7)*, 889-897.

Ferreira, J., Paolella, G., Ramos, C., and Lamond, A.I. (1997). Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories. J Cell Biol *139*, 1597-1610.

Finch, J.T., Lutter, L.C., Rhodes, D., Brown, R.S., Rushton, B., Levitt, M., and Klug, A. (1977). Structure of nucleosome core particles of chromatin. Nature *269*, 29-36.

Finlan, L.E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J.R., and Bickmore, W.A. (2008). Recruitment to the nuclear periphery can alter expression of genes in human cells. PLoS Genet *4*, e1000039.

FitzGerald, K.T., and Diaz, M.O. (1999). MLL2: A new mammalian member of the trx/MLL family of genes. Genomics *59*, 187-192.

French, S.L., Osheim, Y.N., Cioci, F., Nomura, M., and Beyer, A.L. (2003). In exponentially growing *Saccharomyces cerevisiae* cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. Mol Cell Biol *23*, 1558-1568.

Frenster, J.H., Allfrey, V.G., and Mirsky, A.E. (1963). Repressed and active chromatin isolated from interphase lymphocytes. Proc Natl Acad Sci USA *50*, 1026-1032.

Fritze, C.E., Verschueren, K., Strich, R., and Easton Esposito, R. (1997). Direct evidence for SIR2 modulation of chromatin structure in yeast rDNA. EMBO J *16*, 6495-6509.

Fujiki, R., Chikanishi, T., Hashiba, W., Ito, H., Takada, I., Roeder, R.G., Kitagawa, H., and Kato, S. (2009). GlcNAcylation of a histone methyltransferase in retinoic-acid-induced granulopoiesis. Nature *459*, 455-459.

Gartenberg, M.R., Neumann, F.R., Laroche, T., Blaszczyk, M., and Gasser, S.M. (2004). Sir-mediated repression can occur independently of chromosomal and subnuclear contexts. Cell *119*, 955-967.

Geil, C., Schwab, M., and Seufert, W. (2008). A nucleolus-localized activator of Cdc14 phosphatase supports rDNA segregation in yeast mitosis. Curr Biol *18*, 1001-1005.

Goo, Y.H., Sohn, Y.C., Kim, D.H., Kim, S.W., Kang, M.J., Jung, D.J., Kwak, E., Barlev, N.A., Berger, S.L., Chow, V.T., Roeder, R.G., Azorsa, D.O., Meltzer, P.S., Suh, P.G., Song, E.J., Lee, K.J., Lee, Y.C., and Lee, J.W. (2003). Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. Mol Cell Biol *23*, 140-149.

Gotta, M., Laroche, T., Formenton, A., Maillet, L., Scherthan, H., and Gasser, S.M. (1996). The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type *Saccharomyces cerevisiae*. J Cell Biol *134*, 1349-1363.

Gotta, M., Strahl-Bolsinger, S., Renauld, H., Laroche, T., Kennedy, B.K., Grunstein, M., and Gasser, S.M. (1997). Localization of Sir2p: the nucleolus as a compartment for silent information regulators. EMBO J *16*, 3243-3255.

Gottlieb, S., and Esposito, R.E. (1989). A new role for a yeast transcriptional silencer gene, SIR2, in regulation of recombination in ribosomal DNA. Cell *56*, 771-776.

Graumann, J., Dunipace, L.A., Seol, J.H., McDonald, W.H., Yates, J.R., 3rd, Wold, B.J., and Deshaies, R.J. (2004). Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. Mol Cell Proteomics *3*, 226-237.

Gregory, G.D., Vakoc, C.R., Rozovskaia, T., Zheng, X., Patel, S., Nakamura, T., Canaani, E., and Blobel, G.A. (2007). Mammalian ASH1L is a histone methyltransferase that occupies the transcribed region of active genes. Mol Cell Biol *27*, 8466-8479.

Grewal, S.I., and Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. Science *301*, 798-802.

Grummt, I. (1981). Specific transcription of mouse ribosomal DNA in a cell-free system that mimics control in vivo. Proc Natl Acad Sci U S A *78*, 727-731.

Gu, Y., Alder, H., Nakamura, T., Schichman, S.A., Prasad, R., Canaani, O., Saito, H., Croce, C.M., and Canaani, E. (1994). Sequence analysis of the breakpoint cluster region in the ALL-1 gene involved in acute leukemia. Cancer Res *54*, 2327-2330.

Guo, H.B., and Guo, H. (2007). Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity. Proc Natl Acad Sci U S A *104*, 8797-8802.

Haering, C.H., Farcas, A.M., Arumugam, P., Metson, J., and Nasmyth, K. (2008). The cohesin ring concatenates sister DNA molecules. Nature *454*, 297-301.

Halbach, A., Zhang, H., Wengi, A., Jablonska, Z., Gruber, I.M., Halbeisen, R.E., Dehe, P.M., Kemmeren, P., Holstege, F., Geli, V., Gerber, A.P., and Dichtl, B. (2009).

Cotranslational assembly of the yeast SET1C histone methyltransferase complex. EMBO J *28*, 2959-2970.

Hamamoto, R., Furukawa, Y., Morita, M., Iimura, Y., Silva, F.P., Li, M., Yagyu, R., and Nakamura, Y. (2004). SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. Nat Cell Biol *6*, 731-740.

Hamamoto, R., Silva, F.P., Tsuge, M., Nishidate, T., Katagiri, T., Nakamura, Y., and Furukawa, Y. (2006). Enhanced SMYD3 expression is essential for the growth of breast cancer cells. Cancer Sci *97*, 113-118.

Han, Z., Guo, L., Wang, H., Shen, Y., Deng, X.W., and Chai, J. (2006). Structural basis for the specific recognition of methylated histone H3 lysine 4 by the WD-40 protein WDR5. Mol Cell *22*, 137-144.

Harper, D.P., and Aplan, P.D. (2008). Chromosomal rearrangements leading to MLL gene fusions: clinical and biological aspects. Cancer Res *68*, 10024-10027.

Hayashi, K., Yoshida, K., and Matsui, Y. (2005). A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. Nature *438*, 374-378.

Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S.M., and Grunstein, M. (1995). Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. Cell *80*, 583-592.

Hirschhorn, J.N., Bortvin, A.L., Ricupero-Hovasse, S.L., and Winston, F. (1995). A new class of histone H2A mutations in *Saccharomyces cerevisiae* causes specific transcriptional defects in vivo. Mol Cell Biol *15*, 1999-2009.

Hofmann, C., Cheeseman, I.M., Goode, B.L., McDonald, K.L., Barnes, G., and Drubin, D.G. (1998). *Saccharomyces cerevisiae* Duo1p and Dam1p, novel proteins involved in mitotic spindle function. J Cell Biol *143*, 1029-1040.

Holmquist, G.P., and Ashley, T. (2006). Chromosome organization and chromatin modification: influence on genome function and evolution. Cytogenet Genome Res *114*, 96-125.

Hontz, R.D., French, S.L., Oakes, M.L., Tongaonkar, P., Nomura, M., Beyer, A.L., and Smith, J.S. (2008). Transcription of multiple yeast ribosomal DNA genes requires targeting of UAF to the promoter by Uaf30. Mol Cell Biol *28*, 6709-6719.

Hoose, S.A., and Kladde, M.P. (2006). DNA methyltransferase probing of DNA-protein interactions. Methods Mol Biol *338*, 225-244.

Hoppe, G.J., Tanny, J.C., Rudner, A.D., Gerber, S.A., Danaie, S., Gygi, S.P., and Moazed, D. (2002). Steps in assembly of silent chromatin in yeast: Sir3-independent binding of a Sir2/Sir4 complex to silencers and role for Sir2-dependent deacetylation. Mol Cell Biol *22*, 4167-4180.

Hsieh, J.J., Cheng, E.H., and Korsmeyer, S.J. (2003a). Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. Cell *115*, 293-303.

Hsieh, J.J., Ernst, P., Erdjument-Bromage, H., Tempst, P., and Korsmeyer, S.J. (2003b). Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. Mol Cell Biol *23*, 186-194.

Hu, P., Wang, S., and Zhang, Y. (2008). How do SET-domain protein lysine methyltransferases achieve the methylation state specificity? Revisited by Ab initio QM/MM molecular dynamics simulations. J Am Chem Soc *130*, 3806-3813.

Hu, P., and Zhang, Y. (2006). Catalytic mechanism and product specificity of the histone lysine methyltransferase SET7/9: an ab initio QM/MM-FE study with multiple initial structures. J Am Chem Soc *128*, 1272-1278.

Huang, F., Chandrasekharan, M.B., Chen, Y.C., Bhaskara, S., Hiebert, S.W., and Sun, Z.W. (2010). The JmjN domain of Jhd2 is important for its protein stability, and the plant homeodomain (PHD) finger mediates its chromatin association independent of H3K4 methylation. J Biol Chem *285*, 24548-24561.

Huang, J., and Berger, S.L. (2008). The emerging field of dynamic lysine methylation of non-histone proteins. Curr Opin Genet Dev *18*, 152-158.

Huang, J., Brito, I.L., Villen, J., Gygi, S.P., Amon, A., and Moazed, D. (2006a). Inhibition of homologous recombination by a cohesin-associated clamp complex recruited to the rDNA recombination enhancer. Genes Dev *20*, 2887-2901.

Huang, J., and Moazed, D. (2003). Association of the RENT complex with nontranscribed and coding regions of rDNA and a regional requirement for the replication fork block protein Fob1 in rDNA silencing. Genes Dev *17*, 2162-2176.

Huang, J., Perez-Burgos, L., Placek, B.J., Sengupta, R., Richter, M., Dorsey, J.A., Kubicek, S., Opravil, S., Jenuwein, T., and Berger, S.L. (2006b). Repression of p53 activity by Smyd2-mediated methylation. Nature *444*, 629-632.

Huntsman, D.G., Chin, S.F., Muleris, M., Batley, S.J., Collins, V.P., Wiedemann, L.M., Aparicio, S., and Caldas, C. (1999). MLL2, the second human homolog of the Drosophila trithorax gene, maps to 19q13.1 and is amplified in solid tumor cell lines. Oncogene *18*, 7975-7984.

Hyland, E.M., Cosgrove, M.S., Molina, H., Wang, D., Pandey, A., Cottee, R.J., and Boeke, J.D. (2005). Insights into the role of histone H3 and histone H4 core modifiable residues in *Saccharomyces cerevisiae*. Mol Cell Biol *25*, 10060-10070.

Imai, S., Armstrong, C.M., Kaeberlein, M., and Guarente, L. (2000a). Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. Nature *403*, 795-800.

Imai, S., Johnson, F.B., Marciniak, R.A., McVey, M., Park, P.U., and Guarente, L. (2000b). Sir2: an NAD-dependent histone deacetylase that connects chromatin silencing, metabolism, and aging. Cold Spring Harb Symp Quant Biol *65*, 297-302.

Ingvarsdottir, K., Edwards, C., Lee, M.G., Lee, J.S., Schultz, D.C., Shilatifard, A., Shiekhattar, R., and Berger, S.L. (2007). Histone H3 K4 demethylation during activation and attenuation of GAL1 transcription in *Saccharomyces cerevisiae*. Mol Cell Biol *27*, 7856-7864.

Irie, S., Tsujimura, A., Miyagawa, Y., Ueda, T., Matsuoka, Y., Matsui, Y., Okuyama, A., Nishimune, Y., and Tanaka, H. (2009). Single-nucleotide polymorphisms of the PRDM9 (MEISETZ) gene in patients with nonobstructive azoospermia. J Androl *30*, 426-431.

Issaeva, I., Zonis, Y., Rozovskaia, T., Orlovsky, K., Croce, C.M., Nakamura, T., Mazo, A., Eisenbach, L., and Canaani, E. (2007). Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. Mol Cell Biol *27*, 1889-1903.

Jacobs, S.A., Harp, J.M., Devarakonda, S., Kim, Y., Rastinejad, F., and Khorasanizadeh, S. (2002). The active site of the SET domain is constructed on a knot. Nat Struct Biol *9*, 833-838.

Jaehning, J.A. (2010). The Paf1 complex: platform or player in RNA polymerase II transcription? Biochim Biophys Acta *1799*, 379-388.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jessen, W.J., Dhasarathy, A., Hoose, S.A., Carvin, C.D., Risinger, A.L., and Kladde, M.P. (2004). Mapping chromatin structure in vivo using DNA methyltransferases. Methods *33*, 68-80.

Johzuka, K., Terasawa, M., Ogawa, H., Ogawa, T., and Horiuchi, T. (2006). Condensin loaded onto the replication fork barrier site in the rRNA gene repeats during S phase in a FOB1-dependent fashion to prevent contraction of a long repetitive array in *Saccharomyces cerevisiae*. Mol Cell Biol *26*, 2226-2236.

Jones, H.S., Kawauchi, J., Braglia, P., Alen, C.M., Kent, N.A., and Proudfoot, N.J. (2007). RNA polymerase I in yeast transcribes dynamic nucleosomal rDNA. Nat Struct Mol Biol *14*, 123-130.

Jones, R.S., and Gelbart, W.M. (1993). The Drosophila Polycomb-group gene Enhancer of zeste contains a region with sequence similarity to trithorax. Mol Cell Biol *13*, 6357-6366.

Kageyama, S., Liu, H., Kaneko, N., Ooga, M., Nagata, M., and Aoki, F. (2007). Alterations in epigenetic modifications during oocyte growth in mice. Reproduction *133*, 85-94.

Kasahara, K., Ohtsuki, K., Ki, S., Aoyama, K., Takahashi, H., Kobayashi, T., Shirahige, K., and Kokubo, T. (2007). Assembly of regulatory factors on rRNA and ribosomal protein genes in *Saccharomyces cerevisiae*. Mol Cell Biol *27*, 6686-6705.

Kayne, P.S., Kim, U.J., Han, M., Mullen, J.R., Yoshizaki, F., and Grunstein, M. (1988). Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. Cell *55*, 27-39.

Keener, J., Dodd, J.A., Lalo, D., and Nomura, M. (1997). Histones H3 and H4 are components of upstream activation factor required for the high-level transcription of yeast rDNA by RNA polymerase I. Proc Natl Acad Sci U S A *94*, 13458-13462.

Keener, J., Josaitis, C.A., Dodd, J.A., and Nomura, M. (1998). Reconstitution of yeast RNA polymerase I transcription in vitro from purified components. TATA-binding protein is not required for basal transcription. J Biol Chem *273*, 33795-33802.

Keys, D.A., Lee, B.S., Dodd, J.A., Nguyen, T.T., Vu, L., Fantino, E., Burson, L.M., Nogi, Y., and Nomura, M. (1996). Multiprotein transcription factor UAF interacts with the upstream element of the yeast RNA polymerase I promoter and forms a stable preinitiation complex. Genes Dev *10*, 887-903.

Keys, D.A., Vu, L., Steffan, J.S., Dodd, J.A., Yamamoto, R.T., Nogi, Y., and Nomura, M. (1994). RRN6 and RRN7 encode subunits of a multiprotein complex essential for the initiation of rDNA transcription by RNA polymerase I in *Saccharomyces cerevisiae*. Genes Dev *8*, 2349-2362.

Kim, H., Heo, K., Kim, J.H., Kim, K., Choi, J., and An, W. (2009). Requirement of histone methyltransferase SMYD3 for estrogen receptor-mediated transcription. J Biol Chem *284*, 19867-19877.

Kim, T., and Buratowski, S. (2007). Two *Saccharomyces cerevisiae* JmjC domain proteins demethylate histone H3 Lys36 in transcribed regions to promote elongation. J Biol Chem *282*, 20827-20835.

Kladde, M.P., and Simpson, R.T. (1994). Positioned nucleosomes inhibit Dam methylation in vivo. Proc Natl Acad Sci USA *91*, 1361-1365.

Kladde, M.P., and Simpson, R.T. (1996). Chromatin structure mapping in vivo using methyltransferases. Methods Enzymol *274*, 214-233.

Kladde, M.P., Xu, M., and Simpson, R.T. (1996). Direct study of DNA-protein interactions in repressed and active chromatin in living cells. EMBO J *15*, 6290-6300.

Klar, A.J., Strathern, J.N., Broach, J.R., and Hicks, J.B. (1981). Regulation of transcription in expressed and unexpressed mating type cassettes of yeast. Nature *289*, 239-244.

Klein, F., Laroche, T., Cardenas, M.E., Hofmann, J.F., Schweizer, D., and Gasser, S.M. (1992). Localization of RAP1 and topoisomerase II in nuclei and meiotic chromosomes of yeast. J Cell Biol *117*, 935-948.

Klose, R.J., Gardner, K.E., Liang, G., Erdjument-Bromage, H., Tempst, P., and Zhang, Y. (2007). Demethylation of histone H3K36 and H3K9 by Rph1: a vestige of an H3K9 methylation system in *Saccharomyces cerevisiae*? Mol Cell Biol *27*, 3951-3961.

Kobayashi, T. (2003). The replication fork barrier site forms a unique structure with Fob1p and inhibits the replication fork. Mol Cell Biol *23*, 9178-9188.

Kobayashi, T., and Ganley, A.R. (2005). Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. Science *309*, 1581-1584.

Kobayashi, T., Heck, D.J., Nomura, M., and Horiuchi, T. (1998). Expansion and contraction of ribosomal DNA repeats in *Saccharomyces cerevisiae*: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. Genes Dev *12*, 3821-3830.

Kobayashi, T., Hidaka, M., Nishizawa, M., and Horiuchi, T. (1992). Identification of a site required for DNA replication fork blocking activity in the rRNA gene cluster in *Saccharomyces cerevisiae*. Mol Gen Genet *233*, 355-362.

Kobayashi, T., and Horiuchi, T. (1996). A yeast gene product, Fob1 protein, required for both replication fork blocking and recombinational hotspot activities. Genes Cells *1*, 465-474.

Kobayashi, T., Horiuchi, T., Tongaonkar, P., Vu, L., and Nomura, M. (2004). SIR2 regulates recombination between different rDNA repeats, but not recombination within individual rRNA genes in yeast. Cell *117*, 441-453.

Kouskouti, A., Scheer, E., Staub, A., Tora, L., and Talianidis, I. (2004). Gene-specific modulation of TAF10 function by SET9-mediated methylation. Mol Cell *14*, 175-182.

Krajewski, W.A., Nakamura, T., Mazo, A., and Canaani, E. (2005). A motif within SET-domain proteins binds single-stranded nucleic acids and transcribed and supercoiled DNAs and can interfere with assembly of nucleosomes. Mol Cell Biol *25*, 1891-1899.

Kratz, C.P., Emerling, B.M., Donovan, S., Laig-Webster, M., Taylor, B.R., Thompson, P., Jensen, S., Banerjee, A., Bonifas, J., Makalowski, W., Green, E.D., Le Beau, M.M., and Shannon, K.M. (2001). Candidate gene isolation and comparative analysis of a commonly deleted segment of 7q22 implicated in myeloid malignancies. Genomics *77*, 171-180.

Krogan, N.J., Dover, J., Khorrami, S., Greenblatt, J.F., Schneider, J., Johnston, M., and Shilatifard, A. (2002). COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. J Biol Chem *277*, 10753-10755.

Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M., Greenblatt, J.F., and Shilatifard, A. (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol Cell *11*, 721-729.

Kulkens, T., Riggs, D.L., Heck, J.D., Planta, R.J., and Nomura, M. (1991). The yeast RNA polymerase I promoter: ribosomal DNA sequences involved in transcription initiation and complex formation in vitro. Nucleic Acids Res *19*, 5363-5370.

Kunizaki, M., Hamamoto, R., Silva, F.P., Yamaguchi, K., Nagayasu, T., Shibuya, M., Nakamura, Y., and Furukawa, Y. (2007). The lysine 831 of vascular endothelial growth factor receptor 1 is a novel target of methylation by SMYD3. Cancer Res *67*, 10759-10765.

Kurash, J.K., Lei, H., Shen, Q., Marston, W.L., Granda, B.W., Fan, H., Wall, D., Li, E., and Gaudet, F. (2008). Methylation of p53 by Set7/9 mediates p53 acetylation and activity in vivo. Mol Cell *29*, 392-400.

Kwon, T., Chang, J.H., Kwak, E., Lee, C.W., Joachimiak, A., Kim, Y.C., Lee, J., and Cho, Y. (2003). Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. EMBO J *22*, 292-303.

Kyrion, G., Boakye, K.A., and Lustig, A.J. (1992). C-terminal truncation of RAP1 results in the deregulation of telomere size, stability, and function in *Saccharomyces cerevisiae*. Mol Cell Biol *12*, 5159-5173.

Lalo, D., Steffan, J.S., Dodd, J.A., and Nomura, M. (1996). RRN11 encodes the third subunit of the complex containing Rrn6p and Rrn7p that is essential for the initiation of rDNA transcription by yeast RNA polymerase I. J Biol Chem *271*, 21062-21067.

Lappin, T.R., Grier, D.G., Thompson, A., and Halliday, H.L. (2006). HOX genes: seductive science, mysterious mechanisms. Ulster Med J *75*, 23-31.

Laribee, R.N., Shibata, Y., Mersman, D.P., Collins, S.R., Kemmeren, P., Roguev, A., Weissman, J.S., Briggs, S.D., Krogan, N.J., and Strahl, B.D. (2007). CCR4/NOT complex associates with the proteasome and regulates histone methylation. Proc Natl Acad Sci U S A *104*, 5836-5841.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Laroche, T., Martin, S.G., Gotta, M., Gorham, H.C., Pryde, F.E., Louis, E.J., and Gasser, S.M. (1998). Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres. Curr Biol *8*, 653-656.

Le Beau, M.M., Espinosa, R., 3rd, Davis, E.M., Eisenbart, J.D., Larson, R.A., and Green, E.D. (1996). Cytogenetic and molecular delineation of a region of chromosome 7 commonly deleted in malignant myeloid diseases. Blood *88*, 1930-1935.

Lee, J., Saha, P.K., Yang, Q.H., Lee, S., Park, J.Y., Suh, Y., Lee, S.K., Chan, L., Roeder, R.G., and Lee, J.W. (2008a). Targeted inactivation of MLL3 histone H3-Lys-4 methyltransferase activity in the mouse reveals vital roles for MLL3 in adipogenesis. Proc Natl Acad Sci U S A *105*, 19229-19234.

Lee, J.H., and Skalnik, D.G. (2005). CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. J Biol Chem *280*, 41725-41731.

Lee, J.H., and Skalnik, D.G. (2008). Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. Mol Cell Biol *28*, 609-618.

Lee, J.H., Tate, C.M., You, J.S., and Skalnik, D.G. (2007). Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. J Biol Chem *282*, 13419-13428.

Lee, M.G., Wynder, C., Cooch, N., and Shiekhattar, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature *437*, 432-435.

Lee, S., Lee, D.K., Dou, Y., Lee, J., Lee, B., Kwak, E., Kong, Y.Y., Lee, S.K., Roeder, R.G., and Lee, J.W. (2006). Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. Proc Natl Acad Sci U S A *103*, 15392-15397.

Lee, S., Lee, J., Lee, S.K., and Lee, J.W. (2008b). Activating signal cointegrator-2 is an essential adaptor to recruit histone H3 lysine 4 methyltransferases MLL3 and MLL4 to the liver X receptors. Mol Endocrinol *22*, 1312-1319.

Lee, Y., Erkine, A.M., Van Ryk, D.I., and Nazar, R.N. (1995). In vivo analyses of the internal control region in the 5S rRNA gene from *Saccharomyces cerevisiae*. Nucleic Acids Res *23*, 634-640.

Lengronne, A., McIntyre, J., Katou, Y., Kanoh, Y., Hopfner, K.P., Shirahige, K., and Uhlmann, F. (2006). Establishment of sister chromatid cohesion at the *S. cerevisiae* replication fork. Mol Cell *23*, 787-799.

Li, C., Mueller, J.E., and Bryk, M. (2006a). Sir2 represses endogenous polymerase II transcription units in the ribosomal DNA nontranscribed spacer. Mol Biol Cell *17*, 3848-3859.

Li, J., Langst, G., and Grummt, I. (2006b). NoRC-dependent nucleosome positioning silences rRNA genes. EMBO J *25*, 5735-5741.

Li, J., Moazed, D., and Gygi, S.P. (2002). Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. J Biol Chem *277*, 49383-49388.

Li, Y., Reddy, M.A., Miao, F., Shanmugam, N., Yee, J.K., Hawkins, D., Ren, B., and Natarajan, R. (2008). Role of the histone H3 lysine 4 methyltransferase, SET7/9, in the regulation of NF-kappaB-dependent inflammatory genes. Relevance to diabetes and inflammation. J Biol Chem *283*, 26771-26781.

Liang, G., Klose, R.J., Gardner, K.E., and Zhang, Y. (2007). Yeast Jhd2p is a histone H3 Lys4 trimethyl demethylase. Nat Struct Mol Biol *14*, 243-245.

Lin, C.W., Moorefield, B., Payne, J., Aprikian, P., Mitomo, K., and Reeder, R.H. (1996). A novel 66-kilodalton protein complexes with Rrn6, Rrn7, and TATA-binding protein to promote polymerase I transcription initiation in *Saccharomyces cerevisiae*. Mol Cell Biol *16*, 6436-6443.

Linskens, M.H., and Huberman, J.A. (1988). Organization of replication of ribosomal DNA in *Saccharomyces cerevisiae*. Mol Cell Biol *8*, 4927-4935.

Littau, V.C., Allfrey, V.G., Frenster, J.H., and Mirsky, A.E. (1964). Active and inactive regions of nuclear chromatin as revealed by electron microscope autoradiography. Proc Natl Acad Sci USA *52*, 93-100.

Liu, H., Cheng, E.H., and Hsieh, J.J. (2007). Bimodal degradation of MLL by SCFSkp2 and APCCdc20 assures cell cycle execution: a critical regulatory circuit lost in leukemogenic MLL fusions. Genes Dev *21*, 2385-2398.

Liu, J., Wang, X.N., Cheng, F., Liou, Y.C., and Deng, L.W. (2010). Phosphorylation of mixed lineage leukemia 5 by CDC2 affects its cellular distribution and is required for mitotic entry. J Biol Chem *285*, 20904-20914.

Lucchini, R., and Sogo, J.M. (1995). Replication of transcriptionally active chromatin. Nature *374*, 276-280.

Lucchini, R., Wellinger, R.E., and Sogo, J.M. (2001). Nucleosome positioning at the replication fork. EMBO J *20*, 7294-7302.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luna-Fineman, S., Shannon, K.M., and Lange, B.J. (1995). Childhood monosomy 7: epidemiology, biology, and mechanistic implications. Blood *85*, 1985-1999.

Luo, K., Vega-Palas, M.A., and Grunstein, M. (2002). Rap1-Sir4 binding independent of other Sir, yKu, or histone interactions initiates the assembly of telomeric heterochromatin in yeast. Genes Dev *16*, 1528-1539.

Machin, F., Torres-Rosell, J., De Piccoli, G., Carballo, J.A., Cha, R.S., Jarmuz, A., and Aragon, L. (2006). Transcription of ribosomal genes can cause nondisjunction. J Cell Biol *173*, 893-903.

Maillet, L., Boscheron, C., Gotta, M., Marcand, S., Gilson, E., and Gasser, S.M. (1996). Evidence for silencing compartments within the yeast nucleus: a role for telomere proximity and SIR protein concentration in silencer-mediated repression. Genes Dev *10*, 1796-1811.

Malik, S., and Bhaumik, S.R. (2010). Mixed lineage leukemia: histone H3 lysine 4 methyltransferases from yeast to human. FEBS J *277*, 1805-1821.

Mekhail, K., Seebacher, J., Gygi, S.P., and Moazed, D. (2008). Role for perinuclear chromosome tethering in maintenance of genome stability. Nature *456*, 667-670.

Mersman, D.P., Du, H.N., Fingerman, I.M., South, P.F., and Briggs, S.D. (2009). Polyubiquitination of the demethylase Jhd2 controls histone methylation and gene expression. Genes Dev *23*, 951-962.

Merz, K., Hondele, M., Goetze, H., Gmelch, K., Stoeckl, U., and Griesenbeck, J. (2008). Actively transcribed rRNA genes in *S. cerevisiae* are organized in a specialized chromatin associated with the high-mobility group protein Hmo1 and are largely devoid of histone molecules. Genes Dev *22*, 1190-1204.

Metzger, E., Wissmann, M., Yin, N., Muller, J.M., Schneider, R., Peters, A.H., Gunther, T., Buettner, R., and Schule, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature *437*, 436-439.

Milkereit, P., Schultz, P., and Tschochner, H. (1997). Resolution of RNA polymerase I into dimers and monomers and their function in transcription. Biol Chem *378*, 1433-1443.

Milkereit, P., and Tschochner, H. (1998). A specialized form of RNA polymerase I, essential for initiation and growth-dependent regulation of rRNA synthesis, is disrupted during transcription. EMBO J *17*, 3692-3703.

Miller, C.A., and Kowalski, D. (1993). cis-acting components in the replication origin from ribosomal DNA of *Saccharomyces cerevisiae*. Mol Cell Biol *13*, 5360-5369.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., and Shilatifard, A. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc Natl Acad Sci U S A *98*, 12902-12907.

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D., and Hess, J.L. (2002). MLL targets SET domain methyltransferase activity to Hox gene promoters. Mol Cell *10*, 1107-1117.

Milne, T.A., Dou, Y., Martin, M.E., Brock, H.W., Roeder, R.G., and Hess, J.L. (2005). MLL associates specifically with a subset of transcriptionally active target genes. Proc Natl Acad Sci U S A *102*, 14765-14770.

Mishra, B.P., Ansari, K.I., and Mandal, S.S. (2009). Dynamic association of MLL1, H3K4 trimethylation with chromatin and Hox gene expression during the cell cycle. FEBS J *276*, 1629-1640.

Mishra, K., and Shore, D. (1999). Yeast Ku protein plays a direct role in telomeric silencing and counteracts inhibition by rif proteins. Curr Biol *9*, 1123-1126.

Miyamoto, T., Koh, E., Sakugawa, N., Sato, H., Hayashi, H., Namiki, M., and Sengoku, K. (2008). Two single nucleotide polymorphisms in PRDM9 (MEISETZ) gene may be a genetic risk factor for Japanese patients with azoospermia by meiotic arrest. J Assist Reprod Genet *25*, 553-557.

Mo, R., Rao, S.M., and Zhu, Y.J. (2006). Identification of the MLL2 complex as a coactivator for estrogen receptor alpha. J Biol Chem *281*, 15714-15720.

Moazed, D. (2001). Common themes in mechanisms of gene silencing. Mol Cell *8*, 489-498.

Mondoux, M.A., Scaife, J.G., and Zakian, V.A. (2007). Differential nuclear localization does not determine the silencing status of *Saccharomyces cerevisiae* telomeres. Genetics *177*, 2019-2029.

Moretti, P., Freeman, K., Coodly, L., and Shore, D. (1994). Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1. Genes Dev *8*, 2257-2269.

Morris, D.P., Michelotti, G.A., and Schwinn, D.A. (2005). Evidence that phosphorylation of the RNA polymerase II carboxyl-terminal repeats is similar in yeast and humans. J Biol Chem *280*, 31368-31377.

Mueller, J.E., Canze, M., and Bryk, M. (2006). The requirements for COMPASS and Paf1 in transcriptional silencing and methylation of histone H3 in *Saccharomyces cerevisiae*. Genetics *173*, 557-567.

Mulder, K.W., Brenkman, A.B., Inagaki, A., van den Broek, N.J., and Marc Timmers, H.T. (2007). Regulation of histone H3K4 tri-methylation and PAF complex recruitment by the Ccr4-Not complex. Nucleic Acids Res *35*, 2428-2439.

Muller, M., Lucchini, R., and Sogo, J.M. (2000). Replication of yeast rDNA initiates downstream of transcriptionally active genes. Mol Cell *5*, 767-777.

Muntean, A.G., Tan, J., Sitwala, K., Huang, Y., Bronstein, J., Connelly, J.A., Basrur, V., Elenitoba-Johnson, K.S., and Hess, J.L. (2010). The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis. Cancer Cell *17*, 609-621.

Murton, B.L., Chin, W.L., Ponting, C.P., and Itzhaki, L.S. (2010). Characterising the binding specificities of the subunits associated with the KMT2/Set1 histone lysine methyltransferase. J Mol Biol *398*, 481-488.

Musters, W., Knol, J., Maas, P., Dekker, A.F., van Heerikhuizen, H., and Planta, R.J. (1989). Linker scanning of the yeast RNA polymerase I promoter. Nucleic Acids Res *17*, 9661-9678.

Nakamura, T., Blechman, J., Tada, S., Rozovskaia, T., Itoyama, T., Bullrich, F., Mazo, A., Croce, C.M., Geiger, B., and Canaani, E. (2000). huASH1 protein, a putative transcription factor encoded by a human homologue of the Drosophila ash1 gene, localizes to both nuclei and cell-cell tight junctions. Proc Natl Acad Sci U S A *97*, 7284-7289.

Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C.M., and Canaani, E. (2002). ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. Mol Cell *10*, 1119-1128.

Nasmyth, K.A., Tatchell, K., Hall, B.D., Astell, C., and Smith, M. (1981). A position effect in the control of transcription at yeast mating type loci. Nature *289*, 244-250.

Natarajan, T.G., Kallakury, B.V., Sheehan, C.E., Bartlett, M.B., Ganesan, N., Preet, A., Ross, J.S., and Fitzgerald, K.T. (2010). Epigenetic regulator MLL2 shows altered expression in cancer cell lines and tumors from human breast and colon. Cancer Cell Int *10*, 13.

Ng, H.H., Dole, S., and Struhl, K. (2003a). The Rtf1 component of the Paf1 transcriptional elongation complex is required for ubiquitination of histone H2B. J Biol Chem *278*, 33625-33628.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003b). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell *11*, 709-719.

Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C.D., Tempst, P., and Reinberg, D. (2002). Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev *16*, 479-489.

Nislow, C., Ray, E., and Pillus, L. (1997). SET1, a yeast member of the trithorax family, functions in transcriptional silencing and diverse cellular processes. Mol Biol Cell *8*, 2421-2436.

Nogi, Y., Vu, L., and Nomura, M. (1991a). An approach for isolation of mutants defective in 35S ribosomal RNA synthesis in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A *88*, 7026-7030.

Nogi, Y., Yano, R., and Nomura, M. (1991b). Synthesis of large rRNAs by RNA polymerase II in mutants of *Saccharomyces cerevisiae* defective in RNA polymerase I. Proc Natl Acad Sci U S A *88*, 3962-3966.

Odho, Z., Southall, S.M., and Wilson, J.R. (2010). Characterisation of a novel WDR5 binding site that recruits RbBP5 through a conserved motif and enhances methylation of H3K4 by MLL1. J Biol Chem.

Olivares-Illana, V., and Fahraeus, R. (2010). p53 isoforms gain functions. Oncogene *29*, 5113-5119.

Palladino, F., Laroche, T., Gilson, E., Axelrod, A., Pillus, L., and Gasser, S.M. (1993). SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. Cell *75*, 543-555.

Pardo, C., Hoose, S.A., Pondugula, S., and Kladde, M.P. (2009). DNA methyltransferase probing of chromatin structure within populations and on single molecules. Methods Mol Biol *523*, 41-65.

Park, J.H., Cosgrove, M.S., Youngman, E., Wolberger, C., and Boeke, J.D. (2002). A core nucleosome surface crucial for transcriptional silencing. Nat Genet *32*, 273-279.

Park, S., Osmers, U., Raman, G., Schwantes, R.H., Diaz, M.O., and Bushweller, J.H. (2010). The PHD3 domain of MLL acts as a CYP33-regulated switch between MLL-mediated activation and repression. Biochemistry *49*, 6576-6586.

Pasero, P., Bensimon, A., and Schwob, E. (2002). Single-molecule analysis reveals clustering and epigenetic regulation of replication origins at the yeast rDNA locus. Genes Dev *16*, 2479-2484.

Patel, A., Dharmarajan, V., and Cosgrove, M.S. (2008a). Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. J Biol Chem *283*, 32158-32161.

Patel, A., Dharmarajan, V., Vought, V.E., and Cosgrove, M.S. (2009). On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. J Biol Chem *284*, 24242-24256.

Patel, A., Vought, V.E., Dharmarajan, V., and Cosgrove, M.S. (2008b). A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. J Biol Chem *283*, 32162-32175.

Paule, M.R., and White, R.J. (2000). Survey and summary: transcription by RNA polymerases I and III. Nucleic Acids Res *28*, 1283-1298.

Perrod, S., and Gasser, S.M. (2003). Long-range silencing and position effects at telomeres and centromeres: parallels and differences. Cell Mol Life Sci *60*, 2303-2318.

Petes, T.D. (1979). Yeast ribosomal DNA genes are located on chromosome XII. Proc Natl Acad Sci U S A *76*, 410-414.

Petes, T.D., Hereford, L.M., and Botstein, D. (1978a). Simple Mendelian inheritance of the repeating yeast ribosomal DNA genes. Cold Spring Harb Symp Quant Biol *42 Pt 2*, 1201-1207.

Petes, T.D., Hereford, L.M., and Skryabin, K.G. (1978b). Characterization of two types of yeast ribosomal DNA genes. J Bacteriol *134*, 295-305.

Peyroche, G., Milkereit, P., Bischler, N., Tschochner, H., Schultz, P., Sentenac, A., Carles, C., and Riva, M. (2000). The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. EMBO J *19*, 5473-5482.

Pinsky, B.A., Kotwaliwale, C.V., Tatsutani, S.Y., Breed, C.A., and Biggins, S. (2006a). Glc7/protein phosphatase 1 regulatory subunits can oppose the Ipl1/aurora protein kinase by redistributing Glc7. Mol Cell Biol *26*, 2648-2660.

Pinsky, B.A., Kung, C., Shokat, K.M., and Biggins, S. (2006b). The Ipl1-Aurora protein kinase activates the spindle checkpoint by creating unattached kinetochores. Nat Cell Biol *8*, 78-83.

Prasad, R., Yano, T., Sorio, C., Nakamura, T., Rallapalli, R., Gu, Y., Leshkowitz, D., Croce, C.M., and Canaani, E. (1995). Domains with transcriptional regulatory activity within the ALL1 and AF4 proteins involved in acute leukemia. Proc Natl Acad Sci U S A *92*, 12160-12164.

Prasad, R., Zhadanov, A.B., Sedkov, Y., Bullrich, F., Druck, T., Rallapalli, R., Yano, T., Alder, H., Croce, C.M., Huebner, K., Mazo, A., and Canaani, E. (1997). Structure and expression pattern of human ALR, a novel gene with strong homology to ALL-1 involved in acute leukemia and to Drosophila trithorax. Oncogene *15*, 549-560.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods *24*, 218-229.

Quina, A.S., Buschbeck, M., and Di Croce, L. (2006). Chromatin structure and epigenetics. Biochem Pharmacol *72*, 1563-1569.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Reddy, K.L., Zullo, J.M., Bertolino, E., and Singh, H. (2008). Transcriptional repression mediated by repositioning of genes to the nuclear lamina. Nature *452*, 243-247.

Richmond, T.J., and Davey, C.A. (2003). The structure of DNA in the nucleosome core. Nature *423*, 145-150.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., and Klug, A. (1984). Structure of the nucleosome core particle at 7 A resolution. Nature *311*, 532-537.

Rine, J., and Herskowitz, I. (1987). Four genes responsible for a position effect on expression from HML and HMR in *Saccharomyces cerevisiae*. Genetics *116*, 9-22.

Robert, I., Aussems, M., Keutgens, A., Zhang, X., Hennuy, B., Viatour, P., Vanstraelen, G., Merville, M.P., Chapelle, J.P., de Leval, L., Lambert, F., Dejardin, E., Gothot, A., and Chariot, A. (2009). Matrix Metalloproteinase-9 gene induction by a truncated oncogenic NF-kappaB2 protein involves the recruitment of MLL1 and MLL2 H3K4 histone methyltransferase complexes. Oncogene *28*, 1626-1638.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W., Wilm, M., Aasland, R., and Stewart, A.F. (2001). The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. EMBO J *20*, 7137-7148.

Rose, M.D., Winston, F., and Hieter, P. (1990). *Methods in Yeast Genetics: A Laboratory Course Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Rowley, J.D. (1998). The critical role of chromosome translocations in human leukemias. Annu Rev Genet *32*, 495-519.

Ruault, M., Brun, M.E., Ventura, M., Roizes, G., and De Sario, A. (2002). MLL3, a new human member of the TRX/MLL gene family, maps to 7q36, a chromosome region frequently deleted in myeloid leukaemia. Gene *284*, 73-81.

Ruault, M., Dubarry, M., and Taddei, A. (2008). Re-positioning genes to the nuclear envelope in mammalian cells: impact on transcription. Trends Genet *24*, 574-581.

Rusche, L.N., Kirchmaier, A.L., and Rine, J. (2003). The establishment, inheritance, and function of silenced chromatin in *Saccharomyces cerevisiae*. Annu Rev Biochem *72*, 481-516.

Ruthenburg, A.J., Wang, W., Graybosch, D.M., Li, H., Allis, C.D., Patel, D.J., and Verdine, G.L. (2006). Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. Nat Struct Mol Biol *13*, 704-712.

Sadoni, N., Langer, S., Fauth, C., Bernardi, G., Cremer, T., Turner, B.M., and Zink, D. (1999). Nuclear organization of mammalian genomes. Polar chromosome territories build up functionally distinct higher order compartments. J Cell Biol *146*, 1211-1226.

Saffer, L.D., and Miller, O.L., Jr. (1986). Electron microscopic study of *Saccharomyces cerevisiae* rDNA chromatin replication. Mol Cell Biol *6*, 1148-1157.

Santoro, R. (2005). The silence of the ribosomal RNA genes. Cell Mol Life Sci *62*, 2067-2079.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Sayle, R.A., and Milner-White, E.J. (1995). RASMOL: biomolecular graphics for all. Trends Biochem Sci *20*, 374.

Schlichter, A., and Cairns, B.R. (2005). Histone trimethylation by Set1 is coordinated by the RRM, autoinhibitory, and catalytic domains. EMBO J *24*, 1222-1231.

Schnapp, G., Santori, F., Carles, C., Riva, M., and Grummt, I. (1994). The HMG box-containing nucleolar transcription factor UBF interacts with a specific subunit of RNA polymerase I. EMBO J *13*, 190-199.

Schneider, J., Wood, A., Lee, J.S., Schuster, R., Dueker, J., Maguire, C., Swanson, S.K., Florens, L., Washburn, M.P., and Shilatifard, A. (2005). Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. Mol Cell *19*, 849-856.

Schoeftner, S., and Blasco, M.A. (2008). Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. Nat Cell Biol *10*, 228-236.

Seward, D.J., Cubberley, G., Kim, S., Schonewald, M., Zhang, L., Tripet, B., and Bentley, D.L. (2007). Demethylation of trimethylated histone H3 Lys4 in vivo by JARID1 JmjC proteins. Nat Struct Mol Biol *14*, 240-242.

Shahbazian, M.D., and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. Annu Rev Biochem *76*, 75-100.

Shahbazian, M.D., Zhang, K., and Grunstein, M. (2005). Histone H2B ubiquitylation controls processive methylation but not monomethylation by Dot1 and Set1. Mol Cell *19*, 271-277.

Shapiro, R., Braverman, B., Louis, J.B., and Servis, R.E. (1973). Nucleic acid reactivity and conformation. II. Reaction of cytosine and uracil with sodium bisulfite. J Biol Chem *248*, 4060-4064.

Shapiro, R., DiFate, V., and Welcher, M. (1974). Deamination of cytosine derivatives by bisulfite. Mechanism of the reaction. J Am Chem Soc *96*, 906-912.

Shapiro, R., and Weisgras, J.M. (1970). Bisulfite-catalyzed transamination of cytosine and cytidine. Biochem Biophys Res Commun *40*, 839-843.

Shaw, P., and Doonan, J. (2005). The nucleolus. Playing by different rules? Cell Cycle *4*, 102-105.

Shen, L., Guo, Y., Chen, X., Ahmed, S., and Issa, J.P. (2007). Optimizing annealing temperature overcomes bias in bisulfite PCR methylation analysis. Biotechniques *42*, 48, 50, 52 passim.

Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M.R., Lacoste, N., Cayrou, C., Davrazou, F., Saha, A., Cairns, B.R., Ayer, D.E., Kutateladze, T.G., Shi, Y., Cote, J., Chua, K.F., and Gozani, O. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature *442*, 96-99.

Shi, X., Kachirskaia, I., Walter, K.L., Kuo, J.H., Lake, A., Davrazou, F., Chan, S.M., Martin, D.G., Fingerman, I.M., Briggs, S.D., Howe, L., Utz, P.J., Kutateladze, T.G., Lugovskoy, A.A., Bedford, M.T., and Gozani, O. (2007). Proteome-wide analysis in *Saccharomyces cerevisiae* identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. J Biol Chem *282*, 2450-2455.

Shi, Y. (2007). Histone lysine demethylases: emerging roles in development, physiology and disease. Nat Rev Genet *8*, 829-833.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Shi, Y.J., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005). Regulation of LSD1 histone demethylase activity by its associated factors. Mol Cell *19*, 857-864.

Shields, D.J., Altarejos, J.Y., Wang, X., Agellon, L.B., and Vance, D.E. (2003). Molecular dissection of the S-adenosylmethionine-binding site of phosphatidylethanolamine N-methyltransferase. J Biol Chem *278*, 35826-35836.

Shou, W., Sakamoto, K.M., Keener, J., Morimoto, K.W., Traverso, E.E., Azzam, R., Hoppe, G.J., Feldman, R.M., DeModena, J., Moazed, D., Charbonneau, H., Nomura, M., and Deshaies, R.J. (2001). Net1 stimulates RNA polymerase I transcription and regulates nucleolar structure independently of controlling mitotic exit. Mol Cell *8*, 45-55.

Shou, W., Seol, J.H., Shevchenko, A., Baskerville, C., Moazed, D., Chen, Z.W., Jang, J., Charbonneau, H., and Deshaies, R.J. (1999). Exit from mitosis is triggered by Tem1-

dependent release of the protein phosphatase Cdc14 from nucleolar RENT complex. Cell *97*, 233-244.

Siddiqi, I., Keener, J., Vu, L., and Nomura, M. (2001a). Role of TATA binding protein (TBP) in yeast ribosomal dna transcription by RNA polymerase I: defects in the dual functions of transcription factor UAF cannot be suppressed by TBP. Mol Cell Biol *21*, 2292-2297.

Siddiqi, I.N., Dodd, J.A., Vu, L., Eliason, K., Oakes, M.L., Keener, J., Moore, R., Young, M.K., and Nomura, M. (2001b). Transcription of chromosomal rRNA genes by both RNA polymerase I and II in yeast uaf30 mutants lacking the 30 kDa subunit of transcription factor UAF. EMBO J *20*, 4512-4521.

Silva, F.P., Hamamoto, R., Kunizaki, M., Tsuge, M., Nakamura, Y., and Furukawa, Y. (2008). Enhanced methyltransferase activity of SMYD3 by the cleavage of its N-terminal region in human cancer cells. Oncogene *27*, 2686-2692.

Skryabin, K.G., Eldarov, M.A., Larionov, V.L., Bayev, A.A., Klootwijk, J., de Regt, V.C., Veldman, G.M., Planta, R.J., Georgiev, O.I., and Hadjiolov, A.A. (1984). Structure and function of the nontranscribed spacer regions of yeast rDNA. Nucleic Acids Res *12*, 2955-2968.

Skryabin, K.G., Maxam, A.M., Petes, T.D., and Hereford, L. (1978). Location of the 5.8S rRNA gene of *Saccharomyces cerevisiae*. J Bacteriol *134*, 306-309.

Smith, J.S., and Boeke, J.D. (1997). An unusual form of transcriptional silencing in yeast ribosomal DNA. Genes Dev *11*, 241-254.

Song, J.J., and Kingston, R.E. (2008). WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. J Biol Chem *283*, 35258-35264.

Southall, S.M., Wong, P.S., Odho, Z., Roe, S.M., and Wilson, J.R. (2009). Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. Mol Cell *33*, 181-191.

Sperling, A.S., and Grunstein, M. (2009). Histone H3 N-terminus regulates higher order structure of yeast heterochromatin. Proc Natl Acad Sci U S A *106*, 13153-13159.

Stefanovsky, V.Y., Pelletier, G., Bazett-Jones, D.P., Crane-Robinson, C., and Moss, T. (2001). DNA looping in the RNA polymerase I enhancesome is the result of non-cooperative in-phase bending by two UBF molecules. Nucleic Acids Res *29*, 3241-3247.

Steffan, J.S., Keys, D.A., Dodd, J.A., and Nomura, M. (1996). The role of TBP in rDNA transcription by RNA polymerase I in *Saccharomyces cerevisiae*: TBP is required for

upstream activation factor-dependent recruitment of core factor. Genes Dev *10*, 2551-2563.

Stegmeier, F., Visintin, R., and Amon, A. (2002). Separase, polo kinase, the kinetochore protein Slk19, and Spo12 function in a network that controls Cdc14 localization during early anaphase. Cell *108*, 207-220.

Steward, M.M., Lee, J.S., O'Donovan, A., Wyatt, M., Bernstein, B.E., and Shilatifard, A. (2006). Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. Nat Struct Mol Biol *13*, 852-854.

Strahl, B.D., Grant, P.A., Briggs, S.D., Sun, Z.W., Bone, J.R., Caldwell, J.A., Mollah, S., Cook, R.G., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2002). Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Mol Cell Biol *22*, 1298-1306.

Straight, A.F., Shou, W., Dowd, G.J., Turck, C.W., Deshaies, R.J., Johnson, A.D., and Moazed, D. (1999). Net1, a Sir2-associated nucleolar protein required for rDNA silencing and nucleolar integrity. Cell *97*, 245-256.

Stros, M., Launholt, D., and Grasser, K.D. (2007). The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. Cell Mol Life Sci *64*, 2590-2606.

Subramanian, K., Jia, D., Kapoor-Vazirani, P., Powell, D.R., Collins, R.E., Sharma, D., Peng, J., Cheng, X., and Vertino, P.M. (2008). Regulation of estrogen receptor alpha by the SET7 lysine methyltransferase. Mol Cell *30*, 336-347.

Suka, N., Suka, Y., Carmen, A.A., Wu, J., and Grunstein, M. (2001). Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. Mol Cell *8*, 473-479.

Sullivan, M., Higuchi, T., Katis, V.L., and Uhlmann, F. (2004). Cdc14 phosphatase induces rDNA condensation and resolves cohesin-independent cohesion during budding yeast anaphase. Cell *117*, 471-482.

Takahashi, Y.H., Lee, J.S., Swanson, S.K., Saraf, A., Florens, L., Washburn, M.P., Trievel, R.C., and Shilatifard, A. (2009). Regulation of H3K4 trimethylation via Cps40 (Spp1) of COMPASS is monoubiquitination independent: implication for a Phe/Tyr switch by the catalytic domain of Set1. Mol Cell Biol *29*, 3478-3486.

Tan, R.Z., and van Oudenaarden, A. (2010). Transcript counting in single cells reveals dynamics of rDNA transcription. Mol Syst Biol *6*, 358.

Tan, Y.C., and Chow, V.T. (2001). Novel human HALR (MLL3) gene encodes a protein homologous to ALR and to ALL-1 involved in leukemia, and maps to chromosome 7q36 associated with leukemia and developmental defects. Cancer Detect Prev *25*, 454-469.

Tanaka, Y., Katagiri, Z., Kawahashi, K., Kioussis, D., and Kitajima, S. (2007). Trithorax-group protein ASH1 methylates histone H3 lysine 36. Gene *397*, 161-168.

Tanny, J.C., Kirkpatrick, D.S., Gerber, S.A., Gygi, S.P., and Moazed, D. (2004). Budding yeast silencing complexes and regulation of Sir2 activity by protein-protein interactions. Mol Cell Biol *24*, 6931-6946.

Tongaonkar, P., French, S.L., Oakes, M.L., Vu, L., Schneider, D.A., Beyer, A.L., and Nomura, M. (2005). Histones are required for transcription of yeast rRNA genes by RNA polymerase I. Proc Natl Acad Sci U S A *102*, 10129-10134.

Traverso, E.E., Baskerville, C., Liu, Y., Shou, W., James, P., Deshaies, R.J., and Charbonneau, H. (2001). Characterization of the Net1 cell cycle-dependent regulator of the Cdc14 phosphatase from budding yeast. J Biol Chem *276*, 21924-21931.

Tresaugues, L., Dehe, P.M., Guerois, R., Rodriguez-Gil, A., Varlet, I., Salah, P., Pamblanco, M., Luciano, P., Quevillon-Cheruel, S., Sollier, J., Leulliot, N., Couprie, J., Tordera, V., Zinn-Justin, S., Chavez, S., van Tilbeurgh, H., and Geli, V. (2006). Structural characterization of Set1 RNA recognition motifs and their role in histone H3 lysine 4 methylation. J Mol Biol *359*, 1170-1181.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J *13*, 3822-3831.

Tsuge, M., Hamamoto, R., Silva, F.P., Ohnishi, Y., Chayama, K., Kamatani, N., Furukawa, Y., and Nakamura, Y. (2005). A variable number of tandem repeats polymorphism in an E2F-1 binding element in the 5' flanking region of SMYD3 is a risk factor for human cancers. Nat Genet *37*, 1104-1107.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.

Tu, S., Bulloch, E.M., Yang, L., Ren, C., Huang, W.C., Hsu, P.H., Chen, C.H., Liao, C.L., Yu, H.M., Lo, W.S., Freitas, M.A., and Tsai, M.D. (2007). Identification of histone demethylases in *Saccharomyces cerevisiae*. J Biol Chem *282*, 14262-14271.

van Leeuwen, F., Gafken, P.R., and Gottschling, D.E. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell *109*, 745-756.

Verzijlbergen, K.F., Faber, A.W., Stulemeijer, I.J., and van Leeuwen, F. (2009). Multiple histone modifications in euchromatin promote heterochromatin formation by redundant mechanisms in *Saccharomyces cerevisiae*. BMC Mol Biol *10*, 76.

Vidali, G., Gershey, E.L., and Allfrey, V.G. (1968). Chemical studies of histone acetylation. The distribution of epsilon-N-acetyllysine in calf thymus histones. J Biol Chem *243*, 6361-6366.

Vitaliano-Prunier, A., Menant, A., Hobeika, M., Geli, V., Gwizdek, C., and Dargemont, C. (2008). Ubiquitylation of the COMPASS component Swd2 links H2B ubiquitylation to H3K4 trimethylation. Nat Cell Biol *10*, 1365-1371.

Vogelauer, M., Cioci, F., and Camilloni, G. (1998). DNA protein-interactions at the *Saccharomyces cerevisiae* 35 S rRNA promoter and in its surrounding region. J Mol Biol *275*, 197-209.

Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Borchers, C., Tempst, P., and Zhang, Y. (2001). Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. Mol Cell *8*, 1207-1217.

Wang, H.W., Ramey, V.H., Westermann, S., Leschziner, A.E., Welburn, J.P., Nakajima, Y., Drubin, D.G., Barnes, G., and Nogales, E. (2007a). Architecture of the Dam1 kinetochore ring complex and implications for microtubule-driven assembly and force-coupling mechanisms. Nat Struct Mol Biol *14*, 721-726.

Wang, P., Lin, C., Smith, E.R., Guo, H., Sanderson, B.W., Wu, M., Gogol, M., Alexander, T., Seidel, C., Wiedemann, L.M., Ge, K., Krumlauf, R., and Shilatifard, A. (2009). Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. Mol Cell Biol *29*, 6074-6085.

Wang, S., Hu, P., and Zhang, Y. (2007b). Ab initio quantum mechanical/molecular mechanical molecular dynamics simulation of enzyme catalysis: the case of histone lysine methyltransferase SET7/9. J Phys Chem B *111*, 3758-3764.

Wang, S.Z., Luo, X.G., Shen, J., Zou, J.N., Lu, Y.H., and Xi, T. (2008). Knockdown of SMYD3 by RNA interference inhibits cervical carcinoma cell growth and invasion in vitro. BMB Rep *41*, 294-299.

Wang, Z., Song, J., Milne, T.A., Wang, G.G., Li, H., Allis, C.D., and Patel, D.J. (2010). Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression. Cell *141*, 1183-1194.

Waples, W.G., Chahwan, C., Ciechonska, M., and Lavoie, B.D. (2009). Putting the brake on FEAR: Tof2 promotes the biphasic release of Cdc14 phosphatase during mitotic exit. Mol Biol Cell *20*, 245-255.

Warnecke, P.M., Stirzaker, C., Song, J., Grunau, C., Melki, J.R., and Clark, S.J. (2002). Identification and resolution of artifacts in bisulfite sequencing. Methods *27*, 101-107.

Warner, J.R. (1999). The economics of ribosome biosynthesis in yeast. Trends Biochem Sci *24*, 437-440.

Westermann, S., Avila-Sakar, A., Wang, H.W., Niederstrasser, H., Wong, J., Drubin, D.G., Nogales, E., and Barnes, G. (2005). Formation of a dynamic kinetochore-microtubule interface through assembly of the Dam1 ring complex. Mol Cell *17*, 277-290.

Westermann, S., Drubin, D.G., and Barnes, G. (2007). Structures and functions of yeast kinetochore complexes. Annu Rev Biochem *76*, 563-591.

Westermann, S., Wang, H.W., Avila-Sakar, A., Drubin, D.G., Nogales, E., and Barnes, G. (2006). The Dam1 kinetochore ring complex moves processively on depolymerizing microtubule ends. Nature *440*, 565-569.

Wilson, J.R., Jing, C., Walker, P.A., Martin, S.R., Howell, S.A., Blackburn, G.M., Gamblin, S.J., and Xiao, B. (2002). Crystal structure and functional analysis of the histone methyltransferase SET7/9. Cell *111*, 105-115.

Wood, C.M., Nicholson, J.M., Lambert, S.J., Chantalat, L., Reynolds, C.D., and Baldwin, J.P. (2005). High-resolution structure of the native histone octamer. Acta Crystallogr Sect F Struct Biol Cryst Commun *61*, 541-545.

Wu, M., Wang, P.F., Lee, J.S., Martin-Brown, S., Florens, L., Washburn, M., and Shilatifard, A. (2008). Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Mol Cell Biol *28*, 7337-7344.

Wysocka, J., Milne, T.A., and Allis, C.D. (2005a). Taking LSD 1 to a new high. Cell *122*, 654-658.

Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H., and Allis, C.D. (2005b). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell *121*, 859-872.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., Wu, C., and Allis, C.D. (2006). A PHD finger

of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86-90.

Xia, Z.B., Anderson, M., Diaz, M.O., and Zeleznik-Le, N.J. (2003). MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein. Proc Natl Acad Sci U S A *100*, 8342-8347.

Xiao, B., Jing, C., Wilson, J.R., Walker, P.A., Vasisht, N., Kelly, G., Howell, S., Taylor, I.A., Blackburn, G.M., and Gamblin, S.J. (2003). Structure and catalytic mechanism of the human histone methyltransferase SET7/9. Nature *421*, 652-656.

Xiao, L., Williams, A.M., and Grove, A. (2010). The C-terminal domain of yeast high mobility group protein HMO1 mediates lateral protein accretion and in-phase DNA bending. Biochemistry *49*, 4051-4059.

Xu, F., Zhang, Q., Zhang, K., Xie, W., and Grunstein, M. (2007). Sir2 deacetylates histone H3 lysine 56 to regulate telomeric heterochromatin structure in yeast. Mol Cell *27*, 890-900.

Xu, M., Kladde, M.P., Van Etten, J.L., and Simpson, R.T. (1998). Cloning, characterization and expression of the gene coding for a cytosine-5-DNA methyltransferase recognizing GpC. Nucleic Acids Res *26*, 3961-3966.

Yamamoto, R.T., Nogi, Y., Dodd, J.A., and Nomura, M. (1996). RRN3 gene of *Saccharomyces cerevisiae* encodes an essential RNA polymerase I transcription factor which interacts with the polymerase independently of DNA template. EMBO J *15*, 3964-3973.

Yang, M., Gocke, C.B., Luo, X., Borek, D., Tomchick, D.R., Machius, M., Otwinowski, Z., and Yu, H. (2006). Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. Mol Cell *23*, 377-387.

Yang, X.D., Lamb, A., and Chen, L.F. (2009). Methylation, a new epigenetic mark for protein stability. Epigenetics *4*, 429-433.

Yang, X.J. (2004). The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. Nucleic Acids Res *32*, 959-976.

Yokoyama, A., Kitabayashi, I., Ayton, P.M., Cleary, M.L., and Ohki, M. (2002). Leukemia proto-oncoprotein MLL is proteolytically processed into 2 fragments with opposite transcriptional properties. Blood *100*, 3710-3718.

Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D.J., Kitabayashi, I., Herr, W., and Cleary, M.L. (2004). Leukemia proto-oncoprotein MLL forms a SET1-like

histone methyltransferase complex with menin to regulate Hox gene expression. Mol Cell Biol *24*, 5639-5649.

Yu, B.D., Hanson, R.D., Hess, J.L., Horning, S.E., and Korsmeyer, S.J. (1998). MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. Proc Natl Acad Sci USA *95*, 10632-10636.

Zeleznik-Le, N.J., Harden, A.M., and Rowley, J.D. (1994). 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. Proc Natl Acad Sci U S A *91*, 10610-10614.

Zhang, K., Lin, W., Latham, J.A., Riefler, G.M., Schumacher, J.M., Chan, C., Tatchell, K., Hawke, D.H., Kobayashi, R., and Dent, S.Y. (2005). The Set1 methyltransferase opposes Ipl1 aurora kinase functions in chromosome segregation. Cell *122*, 723-734.

Zhang, X., and Bruice, T.C. (2007). Histone lysine methyltransferase SET7/9: formation of a water channel precedes each methyl transfer. Biochemistry *46*, 14838-14844.

Zhang, X., and Bruice, T.C. (2008a). Enzymatic mechanism and product specificity of SET-domain protein lysine methyltransferases. Proc Natl Acad Sci USA *105*, 5728-5732.

Zhang, X., and Bruice, T.C. (2008b). Mechanism of product specificity of AdoMet methylation catalyzed by lysine methyltransferases: transcriptional factor p53 methylation by histone lysine methyltransferase SET7/9. Biochemistry *47*, 2743-2748.

Zhang, X., Yang, Z., Khan, S.I., Horton, J.R., Tamaru, H., Selker, E.U., and Cheng, X. (2003). Structural basis for the product specificity of histone lysine methyltransferases. Mol Cell *12*, 177-185.

Zhou, Z., Ren, X., Huang, X., Lu, L., Xu, M., Yin, L., Li, J., and Sha, J. (2005). SMYD3-NY, a novel SMYD3 mRNA transcript variant, may have a role in human spermatogenesis. Ann Clin Lab Sci *35*, 270-277.

Ziemin-van der Poel, S., McCabe, N.R., Gill, H.J., Espinosa, R., 3rd, Patel, Y., Harden, A., Rubinelli, P., Smith, S.D., LeBeau, M.M., Rowley, J.D., and et al. (1991). Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. Proc Natl Acad Sci USA *88*, 10735-10739.

Zou, J.N., Wang, S.Z., Yang, J.S., Luo, X.G., Xie, J.H., and Xi, T. (2009). Knockdown of SMYD3 by RNA interference down-regulates c-Met expression and inhibits cells migration and invasion induced by HGF. Cancer Lett *280*, 78-85.

APPENDIX A

HUMAN LYSINE 4 HISTONE H3 METHYLTRANSFERASES

A vast amount of literature has been generated in the field of histone methyltransferases with a large focus on understanding histone methyltransferases and their role in leukemia and certain kinds of cancers. Human histone H3 methyltransferases are a complex group, comprising over a hundred different proteins. Specifically there are six known groups of histone methyltransferases that methylate H3K4 in humans: MLL complexes (which include the five separate MLL proteins MLL1-5), Set1A/Set1B complexes, Set7/9, Ash1, Smyd3, and Meisetz. It has been proposed that the need for so many different histone methyltransferases in humans may be due to the need for different HMTs at different developmental stages (Malik and Bhaumik, 2010). In *S. cerevisiae* there is only one SET domain protein responsible for methylation of H3K4. In yeast, Set1 is part of a larger complex called COMPASS (Complex associated with Set1) with seven other subunits that associate with RNA polymerase II (RNA Pol II) to methylate H3K4 (Ng *et al.*, 2003b). Of the human methyltransferases mentioned above, the MLL complexes and human Set1 complexes contain closely related subunits to yeast COMPASS.

SHARED SUBUNITS AMONG HUMAN H3K4 METHYLTRANSFERASES

Certain subunits are commonly shared among SET methyltransferases to form a core complex: WDR5, RBBP5, and Ash2L. The most studied, WDR5 for WD repeat

domain 5, is a doughnut shaped structure with propeller-like domains containing a substrate specific channel through the middle that interacts with dimethylated lysine 4 histone H3 (Wysocka *et al.*, 2005b; Dou *et al.*, 2006). WDR5 contains many copies of a WD40 repeat that is highly conserved from humans to yeast and works as a sort of adaptor protein allowing methyltransferases to interact with a dimethylated histone. WDR5 is a promiscuous subunit of SET domain methyltransferases and interacts with many methyltransferase complexes like the MLL family and human Set1A/Set1B complex. WDR5 is believed to form an initial complex of RBBP5 and Ash2L and then recruit MLL1 for methylation (Steward *et al.*, 2006). A knockdown of WDR5 leads to a decrease in trimethylated H3K4 but methyltransferase complexes are still able to interact with chromatin. WDR5 allows trimethylation of H3K4 and is required for MLL proteins to recognize dimethylated H3K4 (Wysocka *et al.*, 2005b). Crystal structures have clearly shown an N-terminally truncated WDR5 is able to orient and present a tightly bound dimethylated H3K4 substrate to the methyltransferase. Once bound and oriented the dimethylated lysine 4 substrate is stabilized by tyrosine and phenylalanine residues nearby (including Phe133 of WDR5) forming what is typically called in the literature as an 'aromatic cage' (Couture *et al.*, 2006b; Han *et al.*, 2006; Ruthenburg *et al.*, 2006). These crystal structure data are further supported by biochemical data showing WDR5 as a middle man to the histone tail of H3 and the MLL1 methyltransferase complex (Dou *et al.*, 2006). Co-crystallization of WDR5 with a portion of MLL1 identified an arginine containing domain called *Win* for WDR5 interaction motif that is highly conserved among multicellular organisms (Patel *et al.*, 2008a). WDR5 recognizes Arg3765 within

the C-terminal region of SET domain of MLL1, causing the MLL1 complex to form and become enzymatically active (Patel *et al.*, 2008a; Patel *et al.*, 2008b). This interaction between WDR5 and MLL1 occurs within the same surface as interaction with H3K4 and relies on phenylalanine residues (like Phe133) present on the surface of WDR5 (Song and Kingston, 2008). Although most of the data presented here focused on a specific human methyltransferase MLL1, the results are believed to be universal to all human methyltransferase complexes.

Retinoblastoma binding protein 5 (RBBP5) is also a WD40 domain. Not much information exists on the role of RBBP5 in MLL complexes other than it serves as a scaffold protein upon which other members of MLL complexes associate (Dou *et al.*, 2006). As discussed above, WDR5 interacts with MLL1 and the histone H3 tail. On the other side of WDR5 lies a binding site for RBBP5 characterized by a sequence of EVDVT that is conserved in mice and zebra fish but not in *S. cerevisiae* (Odho *et al.*, 2010). RBBP5 is specifically required for interaction with Ash2L (absent, small, or homeotic-like). These two proteins do have a direct effect on H3K4 methylation. With the loss of either RBBP5 or Ash2L, di- and trimethylated H3K4 methylation is decreased (Dou *et al.*, 2006; Steward *et al.*, 2006).

THE MLL FAMILY[4]

The MLL family has five members: MLL1, MLL2, MLL3, MLL4, and MLL5[2]. These members are essential in growth and development and are known for their role in regulating HOX gene expression. HOX proteins have important roles in embryonic development, especially in proper limb development with irregular expression of these HOX proteins common in cancer (Yu *et al.*, 1998; Lappin *et al.*, 2006). The MLL family forms multi-subunit complexes characterized by three core subunits of WDR5, RBBP5, and Ash2L forming a core platform to which other MLL proteins can associate (Dou *et al.*, 2006). These subunits have counterparts in the yeast Set1 COMPASS complex through conserved domains: WDR5 is similar to Swd3, RBBP5 is similar to Swd1, and Ash2L is similar to both Bre2 and Spp1.

The mixed lineage leukemia (MLL) gene, also known as MLL1, ALL-1, HRX, and Htrx, was first discovered due to its link to childhood leukemia such as acute lymphoid leukemia (ALL) and acute myeloid leukemia (AML). A chromosomal translocation occurs within a breakpoint cluster region (BCR) consisting entirely of repeated *Alu* retrotransposons in an intron of the MLL1 gene on chromosome 11 (Ziemin-van der Poel *et al.*, 1991; Djabali *et al.*, 1992; Gu *et al.*, 1994). *Alu* repeats are a hotspot for recombination and are associated with numerous diseases ranging from

[4] Over the years the names of different MLL genes were switched and confused between many papers, making it difficult to present a clear description of each MLL protein. MLL2 and MLL4 were frequently switched in the literature, and MLL5 was also confused with other MLL genes. As a solution the nomenclature used in this dissertation was based upon the National Center for Biotechnology Information (NCBI). Accordingly: MLL1 (ch11q23), MLL2 (ch12q12-q14), MLL3 (ch7q36.1), MLL4 (ch19q13.1), and MLL5 (ch7q22.1). (Ex. **ch**romosome #, then **q** referring to the long arm of the chromosome while **p** refers to the short arm, followed by a numerical designation for location on that arm).

many types of cancers and diabetes to Alzheimer's disease (Deininger and Batzer, 1999). It is estimated that there are ~500,000 copies of the *Alu* repeat in a single human genome, forming 5% of the human genome. Recombination can occur between two *Alu* rich regions or between an *Alu* rich region and a region lacking *Alu* repeats, resulting in a chromosomal translocation where the arms of two different chromosomes are switched. About 80% of childhood leukemias are a direct result of a translocation event within the MLL1 gene that result in the carboxy terminus replaced with over 70 possible different loci (Rowley, 1998; Harper and Aplan, 2008). Initial studies into translocations within MLL1 focused on identifying a specific gene with which translocation occurs; however there does not seem to be one translocation more common than any of the other 70 possible loci.

MLL1 is a protein made up of several different conserved domains including: a cysteine-rich zinc CXXC domain, multiple plant homeodomain (PHD) finger motifs, a bromo domain that recognizes acetylated lysines, a WDR5 interaction or *Win* motif, a nuclear receptor (NR) box and a C-terminal SET domain. MLL1 is a histone methyltransferase with a catalytic SET domain that can add one, two, or three methyl groups to H3K4 using S-adenosyl methionine (Milne *et al.*, 2002; Nakamura *et al.*, 2002). In addition to its ability to methylate, the full MLL1 complex (~1 MDa) plus associated HDACs also acetylate and deacetylate histones (Milne *et al.*, 2002; Nakamura *et al.*, 2002). MLL1 interacts with RNA Pol II transcriptional machinery at the promoter region of a specific set of genes, HOX genes, to regulate transcription (Prasad *et al.*, 1995; Milne *et al.*, 2005).

Initially MLL1 was characterized as having two domains that activate and repress transcription (Zeleznik-Le *et al.*, 1994; Prasad *et al.*, 1995). MLL1 is proteolytically processed at a conserved amino acid sequence to produce an N-terminal (MLL^N/N320/p320) form containing a DNA methyltransferase activity that represses transcription and a C-terminal (MLL^C/C180/p180) form containing the SET domain that activates transcription. These two forms associate together *in vivo* and have different effects on transcription (Nakamura *et al.*, 2002; Yokoyama *et al.*, 2002; Hsieh *et al.*, 2003b). Cleavage occurs by Taspase1 at a conserved amino acid sequence also found in MLL2 (Hsieh *et al.*, 2003a). Once the protein has been cleaved, MLL^C is phosphorylated causing the two forms to associate (Yokoyama *et al.*, 2002) with the remaining MLL1 complex members (Nakamura *et al.*, 2002) and localize together at specific sites within the nucleus (Hsieh *et al.*, 2003b). MLL1 genes that have undergone translocation cannot be cleaved and therefore cannot undergo phosphorylation (Yokoyama *et al.*, 2002; Hsieh *et al.*, 2003b). Cleavage is also required for proper HOX expression, with a decrease in HOX gene expression in cells lacking Taspase1 (Hsieh *et al.*, 2003a). To sum up, a core complex consists of both MLL^N and MLL^C with Ash2L, WDR5, and RBBP5 (Yokoyama *et al.*, 2004). This core complex is generally believed to be conserved among all other MLL complexes. Early work focused on its localization to the promoter region of various HOX genes causing methylation of H3K4 and acetylation of histones (Milne *et al.*, 2002; Nakamura *et al.*, 2002) and activating transcription (Blobel *et al.*, 2009; Robert *et al.*, 2009). The MLL1 complex is recruited to RNA Pol II by the Paf1 complex through MLL1's CXXC domain, leading to

ubiquitylation of histone H2B followed by methylation of H3K4 and initiation of transcription (Chang *et al.*, 2010; Muntean *et al.*, 2010). Studies in mice have shown that MLL1 is believed to provide less than 5% of the total methylated H3K4 present in the cell, specifically at HOX loci, with a loss of MLL1 resulting in decreased H3K4 methylation and decreased RNA Pol II recruitment to these promoter regions (Wang *et al.*, 2009). The majority of methylation seems to be carried out by the human Set1A/Set1B complex (Wu *et al.*, 2008).

The repression domain of MLL1 (MLL^N) contains a DNA methyltransferase domain with a cysteine-rich CXXC region that is critical for binding unmethylated CpG DNA (Birke *et al.*, 2002; Ayton *et al.*, 2004). In humans, DNA methylation is a naturally occurring process that recruits transcriptional silencing complexes forming heterochromatin, while unmethylated regions are associated with areas of active transcription. In *in vitro* assays, binding to specific targets by MLL1 was abolished if DNA is methylated, a sign of heterochromatin (Birke *et al.*, 2002). MLL^N interacts with HDACs through the binding of Cyp33 to the PHD finger domain of MLL1 (Xia *et al.*, 2003). Cyp33 is a protein that binds the PHD zinc finger domain of MLL1 (Fair *et al.*, 2001), causing the function of MLL1 to change from activation of HOX transcription to repression of HOX transcription (Chen *et al.*, 2008; Park *et al.*, 2010).

PHD domains are known for their interactions with methylated lysines residues leading to activation of transcription (Shi *et al.*, 2006; Wysocka *et al.*, 2006), but within MLL1 the PHD domain has a dual role in activation or repression depending upon whether or not Cyp33 associates. The interaction of Cyp33 and the PHD finger of

MLL1 has been studied structurally using NMR spectroscopy. Binding of trimethylated H3K4 by the PHD domain of MLL1 leads to activation of transcription by inhibiting binding of Cyp33, and binding of Cyp33 at the PHD domain of MLL1 inhibits binding of trimethylated H3K4 and represses transcription (Park *et al.*, 2010; Wang *et al.*, 2010). Repression by binding Cyp33 at the PHD domain is characterized by a decrease in the level of methylation of H3K4 and of acetylation of HOX genes although it in unknown how that occurs (Park *et al.*, 2010). The association of Cyp33 binding MLL1 causes the PHD domain, in combination with the bromo domain responsible for recognizing acetylated lysines, to shift structurally in such a way as to inhibit recognition of methylated and acetylated histones, respectively (Wang *et al.*, 2010). Although it is known that binding of Cyp33 at the PHD domain of MLL1 recruits HDACs to MLL1 (Xia *et al.*, 2003), it has not been shown that H3K4 demethylases are recruited.

MLL1 also has a role in methylating and controlling transcription and stability of telomeres in humans (Caslini *et al.*, 2009). MLL1 associates with the nuclear envelope and the outside border of the nucleolus in humans, suggesting a role in the forming heterochromatin (Caslini *et al.*, 2000). MLL1 localizes to telomeres with RNA Pol II to methylate H3K4. Knockdown of MLL1 leads to decreased levels of methylated H3K4 at telomeres, stimulating a DNA damage response and cellular senescence. RNA Pol II transcribes telomeric DNA forming RNA transcripts that are polyadenylated as part of a DNA damage response (Schoeftner and Blasco, 2008; Caslini *et al.*, 2009).

MLL1 is regulated by ubiquitylation and subsequent degradation by the proteasome, and this regulation is controlled at different times within the cell cycle.

MLL1 (both MLL^N and MLL^C) is expressed in two phases throughout the cell cycle, peaking at G1/S and G2/M transitions with degradation by the proteasome in S and M phases by ubiquitylation at the N-terminus (Liu *et al.*, 2007). When the MLL gene undergoes translocation and a fusion protein is made the N-terminus is lost possibly explaining why fusion proteins are so detrimental, i.e. they cannot be degraded properly in response to the cell cycle (Liu *et al.*, 2007). Knockdown of MLL1 leads to decreased proliferation with cell cycle arrest at G2/M phase (Liu *et al.*, 2007; Mishra *et al.*, 2009), but the level of H3K4 methylation remains constant (Mishra *et al.*, 2009). The peak of MLL1 in the cell cycle at G1/S and G2/M coincides with increased levels of RNA Pol II and MLL1 at the promoter region of specific HOX genes resulting in their increased expression (Mishra *et al.*, 2009).

MLL2, also known as ALR (for ALL-1 Related), was originally found due to its similarities to MLL1 with similar PHD domains, CXXC domain and zinc finger motifs, multiple nuclear receptor (NR) boxes, and a SET domain (Prasad *et al.*, 1997). As with MLL1, MLL2 associates with a core complex of ASH2L, RBBP5, and WDR5 (Mo *et al.*, 2006) and associates to the promoter regions of target genes to methylate H3K4 (Issaeva *et al.*, 2007). Cleavage of MLL2 by taspase in humans has been shown, with incomplete cleavage in certain types of cancers (Natarajan *et al.*, 2010).

The location of the MLL2 gene is close to an area known for translocations that result in many types of cancers and tumors (Prasad *et al.*, 1997), however due to changes in its CXXC domain MLL2 very rarely forms fusions (Bach *et al.*, 2009). The few fusion proteins that are formed are not cancer causing. This does not mean that MLL2

does not have a role in cancer. A study of breast and colon cancer cells revealed increased expression of MLL2 (Natarajan *et al.*, 2010). When comparing both CXXC domains of MLL1 and MLL2, both are able to bind DNA however the CXXC domain targets both proteins to different loci (Bach *et al.*, 2009). This difference can be found in changes in the sequence of the CXXC domain of MLL2 (Bach *et al.*, 2009). This supports the idea that the MLL genes are non-redundant and may each have an important role with a subset of particular target genes.

In mice, studies have shown that MLL2 has a role in total transcriptional silencing that occurs during female oocyte development and fertilization (Andreu-Vieyra *et al.*, 2010). Oocytes are transcriptionally active during development until they reach puberty. After this point, oocytes become completely transcriptionally silent until fertilization occurs (Kageyama *et al.*, 2007). A Cre-lox system was utilized in mice to conditionally knockout MLL2. In this system lox sites (recognition sequences for Cre) are inserted flanking a target gene with expression of Cre recombinase to cleave and remove the MLL2 gene when needed. In these mice with a conditional knockout of MLL2, transcriptional silencing does not occur after puberty and there is a loss in di- and trimethylated H3K4 but not in monomethylated H3K4 (Andreu-Vieyra *et al.*, 2010).

As mentioned above MLL proteins contain nuclear receptor interaction motifs (NR boxes) that provide a role in activation of gene expression via nuclear receptors and hormone signaling. Nuclear receptors are factors that can sense hormones and translate these signals into activating gene expression of a target protein. Of the MLL family, MLL1 contains only one NR box while MLL2, 3, and 4 contain three or more NR boxes.

Although MLL1 contains an NR box, no role in nuclear receptor activation has been found. The MLL2 complex, along with the three core members of Ash2L, RBBP5, and WDR5, is known to interact with estrogen receptor α (ERα) to recruit ERα with the MLL2 complex and induce expression of estrogen specific genes (Mo *et al.*, 2006). The association of ERα and the MLL2 complex is dependent on the presence of estrogen, while ChIP experiments showing association of MLL2 to the promoter region of estrogen-inducible genes and a knockdown of MLL2 showing decreased expression of these genes (Mo *et al.*, 2006). This link between MLL2 and ERα supports research showing a link between MLL2 and breast cancer (Natarajan *et al.*, 2010).

In addition to MLL2, both MLL3 and MLL4 interact with nuclear receptors. Much of the nuclear receptor work discussed below has been performed in mice as well as human systems. MLL3 is also known as HALR meaning homologous to ALR and contains many of the same domains as other MLL proteins like PHD domains and a SET domain, as well as multiple nuclear receptor interaction motifs (NR box). However, MLL3 lacks a CXXC domain. Deletions commonly occur within the MLL3 gene resulting in defects in embryonic brain development and myeloid leukemia (Tan and Chow, 2001; Ruault *et al.*, 2002). MLL4 also contains many of the same domains as mentioned above: a CXXC domain, PHD domains, SET domain, and a NR box (FitzGerald and Diaz, 1999; Huntsman *et al.*, 1999). A common activator of nuclear receptors is Asc-2 (for activating signal cointegrator-2) that functions with a variety of nuclear receptors to activate transcription (Goo *et al.*, 2003; Lee *et al.*, 2008b). In mice, Asc-2 together with a MLL protein forms the ASCOM complex: activating signal

cointegrator-2 complex. Asc-2 recruits either MLL3 or MLL4 along with the core components of RBBP5, WDR5, and Ash2l to form ASCOM-MLL3 or ASCOM-MLL4, respectively (Lee *et al.*, 2006; Lee *et al.*, 2008b). Asc-2 is known to recruit these complexes to the liver X receptor that controls cholesterol and fatty acid metabolism to induce expression of target genes and to trimethylate H3K4 (Lee *et al.*, 2008b). Asc-2 also recruits these complexes to the universal retinoic acid receptor (RAR) that is important a variety of areas like growth and development, embryogenesis, and metabolism (Lee *et al.*, 2006). Knockdowns in mice of either MLL3 or MLL4 are still viable with normal levels of gene expression at targeted receptor genes and partial trimethylation suggesting that these two complexes are redundant. Only by affecting both MLL3 and MLL4 is gene expression and trimethylated H3K4 decreased at target genes (Lee *et al.*, 2008a).

Of the five MLL proteins, MLL5 contains only a PHD domain, three nuclear localization signals (NLSs), and a SET domain that is located at the C-terminus and not at the N-terminus like the majority of SET domain proteins. A complete or partial deletion of chromosome 7, the location of the MLL5 gene, is common in myeloid leukemias and is referred to as monosomy 7 (Luna-Fineman *et al.*, 1995; Le Beau *et al.*, 1996; Kratz *et al.*, 2001; Emerling *et al.*, 2002). MLL5 localizes within the nucleus through three nuclear localization signals (NLSs) and forms discrete foci within chromatin. The importance of these foci and NLS domains is unknown (Deng *et al.*, 2004).

MLL5 does have a role in cell cycle progression. Overexpression of MLL5 causes cell cycle arrest at G1, a checkpoint for DNA damage and genome stability (Deng *et al.*, 2004). Knockdown of MLL5 arrests cells in both G1 and G2/M by activating the CDK inhibitor p21, causing downstream targets essential in cell cycle progression to lack phosphorylation required for proper function (Cheng *et al.*, 2008). The protein kinase Cdc2 phosphorylates MLL5 in G2/M phase causing MLL5 to leave chromatin. Once cells exit mitosis MLL5 is dephosphorylated and is able to reassociate with chromatin (Liu *et al.*, 2010). In addition to phosphorylation, MLL5 is modified with β-N-acetylglucosamine (GlcNAc) resulting in the maturation of white blood cells called granulocytes by associating with the nuclear retinoic acid receptor (RARα). Recruitment by these receptors is in concurrence with di- and trimethylated H3K4 at target genes (Fujiki *et al.*, 2009). The methyltransferase activity of MLL5 is dependent upon GlcNAcylation. From this work, MLL5 was shown for the first time to methylate H3K4 however it does not show what if any common MLL subunits might be associated with MLL5.

SET1A AND SET1B METHYLTRANSFERASES

The human Set1A/Set1B complex contains subunits that are homologous to the yeast Set1 COMPASS complex. Set1A was originally found through experiments to identify interacting partners to CFP1, a protein containing a CXXC domain that recognizes unmethylated CpG sites in genomic DNA (Lee and Skalnik, 2005), while Set1B was identified through its similarity to Set1A (Lee *et al.*, 2007). As in yeast,

human Set1 contains an RNA recognition motif (RRM) and a SET domain. Set1A and Set1B associate with the same set of subunits: Ash2L (homolog of Bre2 in yeast), WDR5 (as Swd3 in yeast), RBBP5 (as Swd1 in yeast), CFP1 (as Spp1 in yeast), WDR82/Swd2 (as Swd2 in yeast), and DPY-30 (as Sdc1 in yeast). These subunits share conserved domains and are thought to act accordingly with similar functions and roles in histone methylation. Both Set1A and Set1B form distinct and separate complexes that each bind to a separate regions of euchromatin, shown by confocal microscopy experiments with differential staining between localization of Set1A and Set1B in human nuclei (Lee *et al.*, 2007). It is believed that human Set1A/Set1B complexes are responsible for ~95% of all methylation, with MLL complexes contributing the remaining 5% of total methylation to specifically targeted HOX genes. In mice, knockdown of Set1A and Set1B together causes a slight decrease in the level of trimethylated H3K4 and no effect on the levels of mono- and dimethylated H3K4. The presence of other H3K4 methyltransferases skews the results of these experiments by focusing on the global level of H3K4 methylation (Wu *et al.*, 2008).

Wdr82 is one of the more well studied subunits of human Set1A/Set1B complexes. Wdr82 is a WD40 domain protein similar to WDR5 and a homolog of Swd2 in yeast COMPASS. In yeast, COMPASS interacts with RNA Pol II containing a C-terminal domain (CTD) that is phosphorylated on serine 5 (Ng *et al.*, 2003b). COMPASS interacts with RNA Pol II through the Paf1 complex that is responsible for recruiting Rad6/Bre1 to ubiquitylate lysine 123 of histone H2B, a prerequisite to di- and trimethylation of lysine 4 of histone H3 at active genes (Dover *et al.*, 2002; Ng *et al.*,

2003a; Shahbazian *et al.*, 2005). Human Set1A and Set1B also associate with the phosphorylated serine 5 of RNA Pol II CTD at the promoter region of active genes; however, the Paf1 complex is not used to tether the complex to the polymerase. Instead, human Set1A complex is tethered by Wdr82 by associating with the RRM domain of Set1A and recognizing the phosphorylation of serine 5 of RNA Pol II CTD (Lee and Skalnik, 2008). Knockdown of Wdr82 causes a decrease in Set1A expression, a decrease in the association of the human Set1A complex at targeted promoter regions, and a subsequent decrease in methylated H3K4. While Wdr82 associates with the N-terminal RRM domain of Set1A, the other complex subunits associate with the C-terminus of Set1A near or at the SET domain. By using multiple truncated versions of Flag-tagged Set1A in combination with nuclear extracts, Western analysis of pulldowns show the association of each subunit of the Set1A complex with distinct regions of human Set1A (Lee and Skalnik, 2008). Wdr82 association with chromatin does require histone H2B ubiquitylation by human Bre1 (Wu *et al.*, 2008). Partial knockdown of human Bre1 decreased the association of Wdr82 at actively transcribed genes by approximately 50% and a decrease in total trimethylated H3K4.

SET7/9

Set7 (Wang *et al.*, 2001) and Set9 (Nishioka *et al.*, 2002) were originally published as separate proteins and only later determined to be the same protein. The protein is referred to as Set7/9 in the literature. Set7/9 is a small 366-aa monomer that functions without the commonly associated subunits mentioned above to monomethylate

H3K4 through its catalytic SET domain and activate transcription (Wilson *et al.*, 2002; Kwon *et al.*, 2003; Xiao *et al.*, 2003). For these reasons Set7/9 was the first human H3K4 methyltransferase crystallized (Wilson *et al.*, 2002). Set7/9 has been linked to inflammation of the pancreas and diabetes in mice and humans, specifically by affecting islet β-cells that produce insulin (Li *et al.*, 2008; Deering *et al.*, 2009). Also, Set7/9 is known to modify many non-histone proteins including p53, TBP components TAF10 and TAF7, and estrogen receptor α.

ASH1L

Human Ash1L (ASH1-like) was discovered due to its similarity to Ash1 in *Drosophila melanogaster*. *Drosophila* Ash1 is responsible for regulation of HOX transcription during development. Human Ash1 contains several domains similar to MLL1: a bromo domain known to recognize acetylated lysines, a PHD finger, and a catalytic SET domain (Nakamura *et al.*, 2000). Some controversy surrounds this protein. *In vitro* data has shown Ash1 from *Drosophila* to trimethylate H3K4, H3K9, and H4K20 (Beisel *et al.*, 2002). However, another group reported *in vitro* data with ^{3}H-SAM that human Ash1L mono- and dimethylates H3K36 but not H3K4, H3K9, or H4K20 (Tanaka *et al.*, 2007). Gregory *et al.* 2007 has used ChIP to show that the association of human Ash1L peaks at the promoter regions of actively transcribed genes along with RNA Pol II and trimethylated H3K4, and that Ash1L occupancy is decreased in the middle of the gene where there are high levels of H3K9 and H4K20 methylation (Gregory *et al.*, 2007). Their *in vitro* analysis with purified human Ash1L and ^{3}H-SAM

shows Ash1L methylates only H3K4 and not H3K9 or H4K20. They do not test H3K36 as a possible site of methylation (Gregory *et al.*, 2007). In human and mouse cell lines, Ash1L is a member of general transcription machinery and does not activate transcription at a specific subset of genes (Gregory *et al.*, 2007). Ash1L does associate with HOX genes to trimethylate H3K4, but Ash1L and MLL1 seem to work together at the promoter region of HOX genes. Ash1L and MLL1 associate with the same genes independent of one another and have the same level of occupancy at these genes (Gregory *et al.*, 2007).

SMYD3

Smyd3 is for SET and MYND domain containing 3 where the MYND domain is a zinc finger domain known to bind DNA. Upregulation of Smyd3 is associated with a variety of cancers: colorectal carcinoma, hepatocellular carcinoma, breast cancer, and cervical cancer (Hamamoto *et al.*, 2004; Hamamoto *et al.*, 2006; Wang *et al.*, 2008; Kim *et al.*, 2009; Zou *et al.*, 2009). Within the promoter region of Smyd3, a tandem repeat of CCGCC is an important recognition site for the transcription factor E2F-1 (Tsuge *et al.*, 2005). This repeat, when duplicated in the promoter region causes a higher expression level of Smyd3 that is known to be a common characteristic of the cancers listed above. Smyd3 interacts with RNA Pol II along with RNA helicase (HELZ) and HSP90α (heat shock 90kDa protein α) to di- and trimethylate H3K4 at the promoter regions of selected genes to activate transcription (Hamamoto *et al.*, 2004). RNA helicase is required for

association of Smyd3 to RNA Pol II and the presence of HSP90α is able to enhance methylation of H3K4 *in vitro.*

There are multiple forms of Smyd3, one formed from an internal promoter and another formed by post-translational cleavage. Smyd3-NY is an mRNA transcript variant resulting from an internal promoter within the Smyd3 gene, and is expressed 2.3 fold higher in adult testis than in fetal testis. Smyd3-NY may have a role in spermatogenesis in adult humans (Zhou *et al.*, 2005). A separate form was found from post-translational cleavage of Smyd3 that lacks the first 34 amino acids of the N-terminus (Silva *et al.*, 2008). The N-terminus is responsible for association with HSP90α (Silva *et al.*, 2008). This truncated form has a higher level of total methylated H3K4 than wild type. Cleavage of Smyd3 might be a way to control its histone methyltransferase activity (Silva *et al.*, 2008). It is possible that these two forms could be the same truncated version of Smyd3. Although Silva *et al.* 2008 did not find any other variant of *SMYD3* than full length wild type this data has not been confirmed by others.

Smyd3, like other H3K4 methyltransferases mentioned above, has a role in estrogen receptor mediated transcription. Smyd3 interacts with the ligand binding domain of estrogen receptor α (ERα) (Kim *et al.*, 2009), a nuclear receptor that senses hormones and translates these signals into activating gene expression of a target protein. Smyd3 specifically interacts with ERα to di- and trimethylate H3K4 at the promoter region of specific target genes. With a knockdown of Smyd3, the level of gene

expression of ERα-targeted genes and the level of di- and trimethylated H3K4 decrease (Kim *et al.*, 2009).

MEISETZ

Meisetz was named for its role in controlling transcription during early meiosis: meiosis-induced factor containing a PR/SET domain and zinc-finger motif (Hayashi *et al.*, 2005). The centrally located SET domain of meisetz is unique from all other SET domain proteins in that it lacks the critical domain RXXNHS/C important in catalysis of H3K4 methylation, meaning that this protein catalyzes methylation with a unique mechanism. Meisetz is the first tissue specific histone methyltransferase, located exclusively in fetal ovaries and adult testis with increased expression in cells entering early meiosis during prophase. Meisetz has a role in controlling gametogenesis in both sexes but has been studied more in males. Mice with both copies of meisetz deleted are viable but sterile for both sexes with an accumulation of dimethylated H3K4 in spermatocytes (Hayashi *et al.*, 2005). Meisetz is specific only for dimethylated H3K4 peptide, catalyzing trimethylation. In these cell lines, gametogenesis arrests in both sexes due to mislocalization of DMC1 which is responsible for recognizing double strand breaks. This mislocalization disrupts double strand break repair and damages pairing between homologous chromosomes (Hayashi *et al.*, 2005).

Three single nucleotide polymorphisms identified in meisetz have been linked to azoospermia, a condition affecting ~1% of the male population with low sperm count, infertility, and sterility due to impaired gametogenesis (Miyamoto *et al.*, 2008; Irie *et al.*,

2009). Meisetz has also been linked to recombination. During meiosis, genetic variation is created by maternal and paternal chromosomes assorting randomly forming gametes. Homologous chromosomes can exchange information through recombination termed crossing over at specific "hot spot" regions. Meisetz binds to a 13bp DNA sequence through its zinc finger domain to initiate crossing over at these recombination "hot spots" in the genome (Baudat *et al.*, 2010).

APPENDIX B

HISTONE DEMETHYLASES

Methylation of lysines was once believed to be a permanent mark, with modifications being lost over the course of cell division (Bannister *et al.*, 2002). We now know that there are two groups of histone demethylases: LSD conserved proteins and Jumonji conserved domains. The first demethylase discovered was Human LSD1 for lysine-specific demethylase, which demethylates dimethylated H3K4 (Shi *et al.*, 2004). Sequence analysis of LSD1 found similar proteins conserved in *Arabidopsis, C. elegans, S. pombe,* and *Drosophila;* however, a similar conserved protein in *S. cerevisiae* was not found (Shi *et al.*, 2004). Only through the interaction of LSD1 with a repressor of transcription complex called CoREST can LSD1 demethylate dimethylated H3K4 (Lee *et al.*, 2005; Shi *et al.*, 2005). Crystallization of LSD1 bound to CoREST clearly shows a substrate specific channel for histone H3 (Yang *et al.*, 2006). LSD1's substrate specificity can change to mono- and dimethylated lysine 9 of histone H3 through interacting with the androgen receptor, a nuclear receptor responsible for binding chromatin in the presence of testosterone and activating gene expression (Metzger *et al.*, 2005). Methylated H3K9, a mark present in higher eukaryotes and absent in *S. cerevisiae*, is a mark of heterochromatin and silenced transcription. With this evidence it is unsurprising that elevated levels of LSD1, leading to elevated demethylation of H3K9 and increased gene expression, are a characteristic of prostate cancer (Shi, 2007). To summarize, human LSD1 can function to repress transcription by removing methylation

of H3K4 or can function to activate transcription by removing methylation of lysine 9 of histone H3 depending on what protein complex it is associated with (Wysocka *et al.*, 2005a).

As mentioned above, a conserved LSD1 protein was not found in *S. cerevisiae* (Shi *et al.*, 2004); however four conserved Jumonji (JmjC) domain proteins in yeast have been analyzed: Jhd1, Rph1, Gis1, and Yjr119C/Jhd2/KDM5 (hereafter referred to as Jhd2) (Tu *et al.*, 2007). Jumonji proteins are characterized by a JmjC domain and a PHD domain that associate specifically with methylated tails of histone H3 (Shi *et al.*, 2007). Jhd1 (Tsukada *et al.*, 2006; Fang *et al.*, 2007; Kim and Buratowski, 2007), and Rph1 (Kim and Buratowski, 2007; Klose *et al.*, 2007) both target lysine 36 of histone H3 (methylated by Set2) and characterize transcriptional elongation, while Jhd2 specifically demethylates H3K4 (methylated by Set1 of COMPASS) (Ingvarsdottir *et al.*, 2007; Liang *et al.*, 2007; Seward *et al.*, 2007). Jhd2 utilizes its PHD domain to associate with active and silent chromatin to regulate H3K4 methylation (Huang *et al.*, 2010). Deletion of Jhd2 causes an increase in methylated histones and an increase in Set1 at regions of active transcription (Ingvarsdottir *et al.*, 2007; Liang *et al.*, 2007; Huang *et al.*, 2010). H3K4 methylation has been shown to depend on specific subunits of the proteasome (Laribee *et al.*, 2007; Mulder *et al.*, 2007), but it wasn't until the discovery that Jhd2 is polyubiquitinated by Not4, a ubiquitin ligase, and degraded by the proteasome (Mersman *et al.*, 2009) that it was suggested that Not4 could control demethylation by Jhd2 to ensure proper levels of methylated H3K4 (Huang *et al.*, 2010).

APPENDIX C

ALTERNATIVE SUBSTRATES OF HUMAN LYSINE 4 HISTONE H3 METHYLTRANSFERASES

Human SET domain proteins have been found to methylate non-histone proteins at lysine residues forming new roles of SET domain proteins (Huang and Berger, 2008). Human H3K4 methyltransferases have seven known non-histone substrates: p53, TAF10 and TAF7, estrogen receptor α, NF-κB, Dnmt1, and VEGFR1. There is not a consensus sequence when comparing the amino acid sequences of these proteins with the N-terminus sequence of histone H3 (Yang *et al.*, 2009). NF-κB, a transcription factor involved in inflammatory and immune responses, and Dnmt1, a DNA methyltransferase, are monomethylated by Set7/9. Methylation of these targets leads to increased degradation by the proteasome and are discussed elsewhere (Yang *et al.*, 2009). VEGFR1, for vascular endothelial growth factor receptor 1, is tyrosine kinase receptor with a role in angiogenesis and vascular development is dimethylated by Smyd3 causing elevated autophosphorylation. This topic is not a focus of this dissertation and is discussed elsewhere (Kunizaki *et al.*, 2007). Although these are alternative substrates, the interaction and the role of certain amino acids are similar to when Set7/9 interacts with lysine 4 of histone H3. In some ways, these studies have led to a greater understanding of certain amino acids within Set7/9 than studies with histone H3. These studies have also broadened the understanding of the role of Set7/9 in cancer outside of its role in methylating H3K4.

P53 TUMOR SUPPRESSOR

p53 is a DNA binding protein that is highly expressed in tumor cells in response to stress or damage, leading to a change in gene expression of many target gene. p53 is mutated in 50% of human cancers. p53 expression has been linked to apoptosis, cell cycle arrest, senescence, DNA repair and metabolism, forming the conclusion that p53 functions to inhibit growth and invasion of tumor cells (Olivares-Illana and Fahraeus, 2010). p53 is a lysine-rich protein known to be modified by many post-translational modifications such as phosphorylation, ubiquitylation, acetylation, and sumoylation. Set7/9 monomethylates p53 at K372 in a region of the N-terminus that contains the additional modifications listed above (Chuikov *et al.*, 2004). Monomethylation by Set7/9 of p53 leads to increased expression of its target genes (Chuikov *et al.*, 2004) and, in mice, methylation at K372 is required for acetylation of p53 (Kurash *et al.*, 2008). While methylation by Set7/9 activates p53, methylation by Smyd2 represses p53. Smyd2 is a lysine methyltransferase that methylates lysine 36 of histone H3, not to be confused with Smyd3 that methylates lysine 4 of histone H3, that monomethylates p53 at K360 and represses the binding of p53 to the promoter regions of target genes and the ability of p53 to activate target genes. Methylation of p53 at K372 by Set7/9 inhibits methylation at K370 by Smyd2 (Huang *et al.*, 2006b).

Structural studies of Set7/9 with a peptide of p53 shows positioning of the target lysine of the p53 peptide within the same substrate specific channel as the H3 tail described above, however interactions with active site amino acids are different. Five amino acids within Set7/9 form important interactions with the p53 peptide, but only

Tyr335 has a dual role in p53 and the histone H3 tail (Chuikov *et al.*, 2004). In interacting with p53, the position of Tyr335 is highly similar to its position in relation to lysine 4 of histone H3, forming hydrogen bonds to the backbone of the p53 peptide. Computer simulations identical to that described above were performed with Set7/9 bound with a p53 peptide and s-adenosyl homocysteine, and produced similar results as with lysine 4 of histone H3. A water channel is required for deprotonation of lysine 372 of p53 leading to monomethylation through a SN_2 reaction. Positioning of monomethylated p53 prevents the formation of a second water channel and further methylation (Zhang and Bruice, 2008b).

TAF10 AND TAF7

TAF10 and TAF7 are components of the TFIID transcription factor complex that binds within the promoter region of genes to recruit RNA Pol II. Set7/9 monomethylates TAF10 at K189 *in vivo*, increasing the association of TAF10 with RNA Pol II and thus increasing the transcription levels of specific genes (Kouskouti *et al.*, 2004). Set7/9 methylates TAF7 *in vitro* at K5 but this has not been supported by *in vivo* analysis and this effect of this methylation on RNA Pol II is unknown (Couture *et al.*, 2006a).

Structural studies of Set7/9 associated with TAF10 show Tyr335 as the only amino acid in common with Set7/9 interacting with lysine 4 of histone H3. As before, Tyr335 forms hydrogen bonds with the backbone of its substrate TAF10 backbone (Couture *et al.*, 2006a). Crystal structures of Set7/9 mutants Y245A and Y305F with TAF10 peptides provided similar results discussed above with crystal structures of wild

type Set7/9 with lysine 4 of histone H3 (Del Rizzo *et al.*, 2010). The two mutated Set7/9 structures differ between 0.3-0.4 Å from the wild type structure and thus provide information of how these mutations affect the function of Set7/9. As with lysine 4 of histone H3, Y305F allows dimethylation of TAF10 while Y245A allows trimethylation of TAF10 by Set7/9 (Del Rizzo *et al.*, 2010). Y305F binds unmodified and monomethylated TAF10 peptides 4-7 fold more tightly than wild type, while Y245A binds unmodified and methylated substrates with equal affinities. In the Y305F mutant, Y245 is responsible for positioning the lysine residue of TAF10 through hydrogen bonding with its hydroxyl group. In the Y245A mutant, Y305 is responsible for positioning the lysine residue through hydrogen bonds to mono- and dimethylated TAF10. Positioning of water molecules and hydrogen bonding has changed within the active site in both mutants, but nothing is mentioned on how these mutants affect the formation of a water channel. Both Y305F and Y245A follow a distributive model of methylation, where intermediate products are released by Set7/9 before further methylation occurs (Del Rizzo *et al.*, 2010). This conflicts with wild type data that shows forming monomethylated H3K4 is distributive, while di- to trimethylation is successive (Dirk *et al.*, 2007).

ESTROGEN RECEPTOR α

Estrogen receptor α (ERα) is a nuclear receptor known to interact with a core complex of MLL2, Ash2L, RBBP5, and WDR5 in the presence of estrogen to induce expression of estrogen-specific genes (Mo *et al.*, 2006). ERα is monomethylated at

K302 by Set7/9. Knockdown of Set7/9 results in a decrease in protein levels of ERα, possibly through proteasome degradation, and a decrease in activation of transcription of estrogen-specific genes (Subramanian *et al.*, 2008). Methylation of ERα by SET7/9 is required for localization of ERα to the promoter region of target genes. A crystal structure was created with Set7/9, an ERα peptide, and S-adenosyl homocysteine. As with other structures, the targeted lysine 302 lies within a narrow substrate channel with Tyr305 pointing towards the lysine (Subramanian *et al.*, 2008).

VITA

NAME	Kelly M. Williamson
ADDRESS	Department of Biochemistry and Biophysics Texas A&M University College Station, Texas 77845-2128
E-MAIL ADDRESS	kell101@gmail.com
EDUCATION	B.S. Biochemistry (2004) Texas State University San Marcos, Texas 78666
	Ph.D. Biochemistry (2011) Texas A&M University College Station, Texas 77845-2128